HANDBOOK OF

Microlithography, Micromachining, and Microfabrication

Volume 2: MICROMACHINING AND MICROFABRICATION

Editor
P. RAI-CHOUDHURY

IEE MATERIALS AND DEVICES SERIES 12B

Digitized by the Internet Archive
in 2022 with funding from
Kahle/Austin Foundation

https://archive.org/details/handbookofmicrom0002prai

intensephotonics

Intense Photonics
Block 7 Kelvin Campus
West of Scotland Science Park
Glasgow G20 0TH
Scotland UK

Tel +44 (0)141 579 3030
Fax +44 (0)141 579 3044
Email@intensephotonics.com
www.intensephotonics.com

HANDBOOK OF

Microlithography, Micromachining, and Microfabrication

Volume 2: MICROMACHINING AND MICROFABRICATION

EDITORIAL BOARD

A. Bruno Frazier
University of Utah

Patrick J. French
Delft University of Technology

James W. Knutti
Exar Corporation

Paul J. McWhorter
Sandia National Labs

Dennis L. Polla
University of Minnesota

Ray M. Roop
Motorola

Osamu Tabata
Ritsumeikan University

Vijay Varadan
The Pennsylvania State University

SPIE PRESS MONOGRAPH PM40

IEE MATERIALS AND DEVICES SERIES 12B
Series Editors: Professor A. J. Moses
Dr. John Wood

HANDBOOK OF

Microlithography, Micromachining, and Microfabrication

Volume 2: MICROMACHINING AND MICROFABRICATION

P. Rai-Choudhury, *Editor*

SPIE Optical Engineering Press
A Publication of SPIE—The International Society for Optical Engineering
Bellingham, Washington USA

IEE
The Institution of Electrical Engineers
London, UK

Library of Congress Cataloging-in-Publication Data

Handbook of microlithography, micromachining, & microfabrication / P. Rai
-Choudhury, editor.
p. cm.
Includes bibliographical references and index.
ISBN 0-8194-2379-3 (v. 2)
1. Microlithography. 2. Micromachining. 3. Microfabrication.
I. Rai-Choudhury, P.
TK7836.H3423 1997
670—dc21 96-40237
CIP

Copublished by

SPIE—The International Society for Optical Engineering
P.O. Box 10
Bellingham, Washington 98227-0010
Phone: 360/676-3290
Fax: 360/647-1445
Email: spie@spie.org
WWW: http://www.spie.org/
SPIE Press Monograph PM40: ISBN 0-8194-2379-3

The Institution of Electrical Engineers
Michael Faraday House
Six Hills Way, Stevenage, Herts.
SG1 2AY United Kingdom
Phone: +44 (0)1438 313311
Fax: +44 (0)1438 360079
Email: books@iee.org.uk
WWW: http://www.iee.org.uk
IEE Materials and Devices Series 12B: ISBN 0-85296-911-2

Copyright © 1997 The Society of Photo-Optical Instrumentation Engineers

All rights reserved. No part of this publication may be reproduced or distributed in any form or by any means without written permission of the publisher.

Printed in the United States of America.

CONTENTS

PREFACE vii

1 MICROMACHINING AND TRENDS FOR THE TWENTY-FIRST CENTURY 3
Paul J. McWhorter, A. Bruno Frazier, P. Rai-Choudhury

2 WET CHEMICAL ETCHING OF SILICON AND SiO_2, AND TEN CHALLENGES FOR MICROMACHINERS 41
Don L. Kendall and Robert A. Shoultz

3 APPLICATIONS OF DRY ETCHING TO MICROSENSORS, FIELD EMITTERS, AND OPTICAL DEVICES 99
Stella W. Pang

4 FOCUSED ION BEAMS FOR MICROMACHINING AND MICROCHEMISTRY 153
Diane K. Stewart and J. David Casey, Jr.

5 PLATING TECHNIQUES 197
Lubomyr T. Romankiw and Eugene J. M. O'Sullivan

6 HIGH ASPECT RATIO PROCESSING 299
Craig R. Friedrich, Robert Warrington, Walter Bacher, Werner Bauer, Philip J. Coane, Jost Göttert, Thomas Hanemann, Jürgen Haußelt, Mathias Heckele, Regina Knitter, Jürgen Mohr, Volker Piotter, Hans-Joachim Ritzhaupt-Kleissl, Robert Ruprecht

7 SENSORS AND ACTUATORS 379
Hal Jerman and Stephen Terry

8 MICRO-OPTICAL DEVICES 435
Hiroyuki Fujita, Hiroshi Toshiyoshi

9 MICROMACHINED SYSTEMS FOR NEUROPHYSIOLOGICAL APPLICATIONS 517
Khalil Najafi

10 THIN-FILM-TRANSISTOR-ADDRESSED LIQUID CRYSTAL DISPLAYS 571
Fang-Chen Luo

11 MICROSENSORS, ACTUATORS, MEMS, AND ELECTRONICS FOR SMART STRUCTURES 617
Vijay K. Varadan and Vasundara V. Varadan

INDEX 689

PREFACE

Microlithography, micromachining and microfabrication are rapidly finding applications in many areas, from sensors and actuators to biomedical devices, as well as their uses in microelectronics device manufacturing. Lithography is the key technology that has driven the dynamic growth of the IC industry over the past two decades. To date, optical lithography continues to be the mainstream technology for the IC industry, and is being used in production by leading-edge high-volume manufacturers to support 0.25-μm minimum feature size. Although the exposure system using 193-nm optical lithography is expected to extend to 0.13 μm, the industry remains undecided as to the choice of an exposure system beyond 0.13 μm. The options include extreme ultraviolet (EUV or projection x-ray), e-beam projection, massive parallel direct write, and 1x proximity x-ray. The field of lithography will continue to be very dynamic, and demands an authoritative handbook for process development and production to aid in the training of scientists and engineers.

Microlithography and micromachining are also driving microelectromechanical systems (MEMS). The MEMS technology is rapidly spreading internationally. Within the next decade the cost of micromachined devices will drop to the point where there will be explosive demand for such devices for use in such industries as automotive, aircraft, and disposable medical products. In addition, MEMS will find applications for in-situ process monitoring, environmental health and safety monitoring, and numerous other areas. Use of lithography for fabrication of many microelectromechanical devices frequently requires processing procedures that range from the fabrication of high-aspect-ratio to ultrafine structures.

There are a number of books on lithography, but there still is a need to compile all the diverse information into an easily accessible handbook-type format. SPIE PRESS is publishing the *Handbook of Microlithography, Micromachining, and*

Microfabrication in two volumes. Volume 1 is on Microlithography, and Volume 2 is on Micromachining and Microfabrication.

Volume 2 focuses on the process technology and the numerous device applications using micromachining and microfabrication. This handbook will be useful to researchers and engineers who are not specialists in the field and will serve as a reference book for the experts. It will also attempt to collate pertinent available resource information. The material is presented in such a manner as to allow the book to be utilized as both a textbook and reference book. This handbook is intended for use primarily in industrial laboratories, universities, libraries, and manufacturing facilities.

Acknowledgments

Special thanks must go to my wife, Margaret, for all of her expertise with the English language and for her support thoughout this project. I would like to thank Mary Barnard, Susan Price, and Ruth Haas for their editing and organization of the book, and Eric Pepper for encouragement to undertake this handbook.

P. Rai-Choudhury
August 1997

HANDBOOK OF

Microlithography, Micromachining, and Microfabrication

Volume 2: MICROMACHINING AND MICROFABRICATION

CHAPTER 1

Micromachining and Trends for the Twenty-First Century

Paul J. McWhorter
Sandia National Laboratories

A. Bruno Frazier
University of Utah

P. Rai-Choudhury
SPIE—The International Society for Optical Engineering

CONTENTS

1.1 Introduction / 4

1.2 Application Examples / 10
  1.2.1 Displays / 10
  1.2.2 Inertial sensors / 12
  1.2.3 HP ink jet print head / 14
  1.2.4 Biomedical instrumentation / 14
    1.2.4.1 Neural stimulation/recording instrumentation / 15
    1.2.4.2 Biological/chemical analysis systems / 16
    1.2.4.3 Whole cell systems / 18
    1.2.4.4 Precision surgical tools and medical monitoring instrumentation / 19
    1.2.4.5 Prosthetic devices and tactile sensors / 20

1.3 Fabrication Technologies / 20
  1.3.1 Bulk micromachining / 20
  1.3.2 Surface micromachining / 21
  1.3.3 Mold micromachining / 21

1.4 Motivation To Integrate MEMS and Microelectronics / 22
  1.4.1 Integration strategies / 23
  1.4.2 Microelectronics first, micromachines last / 24
  1.4.3 Micromachines in the middle / 24
  1.4.4 Micromachines first / 27
  1.4.5 Considerations in determining integration strategies / 29

1.5 Future Potential / 30

References / 30

1.1 INTRODUCTION

Imagine a machine so small that it is imperceptible to the human eye. Imagine working machines with gears no bigger than a grain of pollen, dwarfed in size by a microscopic dust mite (Fig. 1.1). Imagine these machines being batch fabricated tens of thousand at a time, at a cost that begins to approach zero. Imagine a realm where the world of design is turned upside down, and the seemingly impossible suddenly becomes easy—a place where gravity and inertia are no longer important, but the effects of atomic forces and surface science dominate. Welcome to the microdomain, a world now occupied by an explosive new technology known as MEMS (MicroElectroMechanical Systems) or, more simply, micromachines.

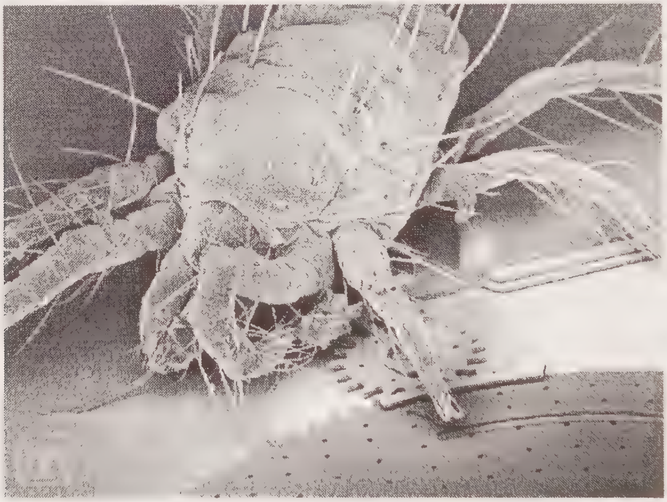


FIG. 1.1 Sandia National Laboratories' MEMS locking mechanism is dwarfed in size by a microscopic dust mite.

MEMS is the next logical step in the silicon revolution. The silicon revolution began over three decades ago, with the introduction of the first integrated circuit.[1] The integrated circuit has changed virtually every aspect of our lives and has been the key defining technology of the twentieth century. The hallmark of the integrated circuit industry over the past three decades has been the exponential increase in the number of transistors incorporated onto a single piece of silicon. This rapid advance in the number of transistors per chip leads to integrated circuits with continuously increasing capability and performance. Continual advances in the manufacturing process control, cleanliness, and critical dimension precision have resulted in the ability to make devices in which the cost of individual transistors or memory cells becomes essentially free. Integrated circuits, which contain tens of millions of transistors, all virtually identical, and all operating as expected, are now routinely made at an affordable cost. As the cost of the basic integrated circuit building block becomes essentially free, the computation, processing, and communication power that can be achieved in an affordable desktop computer staggers the imagination. Processing power that just a few years ago could only be found on the most expensive computers at the world's best research facilities can now be readily acquired in the average household. With this affordable computation power, integrated circuits have become ubiquitous. They are becoming part of virtually all commercial, consumer, and military products. As time has progressed, large, expensive, complex systems have been replaced by small, high-performance, inexpensive integrated circuits.

While the growth in the functionality of microelectronic circuits has been truly phenomenal, for the most part this growth has been limited to data processing, storage, and communication. It is interesting to compare the desktop computer of today with the computer of ten years ago. In many ways, the computer of today is very different than the computer of a decade ago, but in other ways, they are very similar. While the raw processing power, speed, and data storage of today's computers is light years ahead of a decade ago, the two computers look basically the same. In both cases we see people sitting in front of boxes, providing information to the box through the keyboard, and getting information from the box via a CRT. While what the box can do has seen dramatic advances, the way we interact with the box remains basically unchanged. This is because the advancement of the integrated circuit technology has been driven solely by the objective of shrinking the linear dimension of the transistor, the basic building block of the integrated circuit. As the linear dimension of the transistor shrinks, it gets faster, and you can get more of them on a given piece of silicon. This results in chips that can process and store more information faster. To date, this has been the silicon revolution.

The next silicon revolution will be different and more important than simply packing more transistors onto the silicon. The power and capability of the chips of tomorrow will transcend simply storing and processing information. We believe that the hallmark of the next 30 years of the silicon revolution will be the incorporation of new types of functionality onto the chip; structures that will enable the chip to not only reason, but to sense, act, and communicate as well. This will result in chips that will have an ability to affect their surroundings and to communicate what they know to the outside world in new and dramatic ways. Besides processing data, the chips of tomorrow will process such things as chemicals, motion, light, and knowledge. This will be the hallmark of the next silicon revolution, and will have as profound an effect on the next century as the integrated circuit has had on the last.

With the next silicon revolution, the metrics to mark our progress will change. To date, the technology has measured its march by a single metric: the number and size of transistors that can be packed on a single piece of silicon. This metric will be supplanted by metrics that measure how capable a chip is, i.e., to what degree the chip can understand its environment, affect its environment, and enlighten its user.

Dramatic advances in the MEMS field over the last five years begin to show a glimpse into the realization of the new silicon revolution. MEMS enable the fabrication of machines that are vanishingly small. These machines have the ability to perform a variety of functions, including: physical[1-9] and chemical sensing,[10-14] actuation,[15-19] steering light,[20-24] and communication.[25-27] This field has generated intense interest and widespread speculation about the future applications of these minuscule machines. Much of the speculation has centered around the most obvious characteristic of these machines: their size. The fact that these machines are small has captured the imagination of the world. People have been speculating about free-ranging, autonomous robots cruising through the microdomain performing useful work: cleaning our blood veins, repairing broken nerves, repairing tiny defects in integrated circuits, and so forth. For the most part, much of this speculation has missed the mark because it has focused on the most obvious, and not the most important characteristic of MEMS. The most obvious characteristic of MEMS is that they are small. The most important characteristic of MEMS is that they can be made cheaply. To explain this, consider the advanced Pentium processor in a desktop computer. Let's assume for a second that the Pentium chip itself occupied a volume of 100 times its present size, with its other performance and price characteristics the same. The overall size, weight, and form of the desktop computer would not be affected. In fact, this is true of most all applications of integrated circuits. The overall impact of integrated circuits on the world around us would not be dramatically affected if they were much larger. On the other hand, let's assume for a moment that Pentium processors were 100X as expensive. Now, all of the sudden, many applications where they are presently used could no longer afford them. Their overall impact on the world would be significantly less. Hence, the real impact of integrated circuits, and analogously MEMS, does not come from their size, but from their cost.

MEMS can be made cheaply because they are made by exploiting the existing integrated circuit manufacturing infrastructure.[28-30] Integrated circuits are made by the successive deposition, photopatterning, and then etching of thin films on silicon.[31-32] For the case of integrated circuits, these patterns are formed to create small electrical devices. For the case of MEMS, these same fabrication sequences are used to create mechanical structures. Over the past 30 years, there has been tremendous investment in the microelectronics infrastructure worldwide. The ability of the MEMS field to exploit this existing infrastructure will provide huge momentum in the continued development of this field. In addition to propelling the MEMS field forward, this synergy between MEMS and ICs will have ancillary benefits to the integrated circuit industry. Due to the rapid advance of the integrated circuit industry, microelectronic calibration equipment can become obsolete in a very small number of years. As the circuit complexities continue to increase, microelectronic fabrication facilities quickly end up with no commercially viable circuits that can be produced in them. The primary reason for this is associated with the continued advancement in the requirements on the photolithography and associated processes needed to create ever smaller critical dimensions. If we look at the photolithography requirements for MEMS, however, few applications require linewidths of less than one micron, so a fabrication facility that is obsolete from the perspective of integrated circuits could be state of the art from the perspective of MEMS. Thus, MEMS offers the opportunity to extend the useful life of aging microelectronic fabrication facilities.

As with any great endeavor, the future of MEMS will have winners and it will have losers. MEMS has come to mean many different things, and there is no widely accepted definition of what is and is not MEMS. To some, MEMS is only those technologies that are made using completely standard integrated circuit fabrication processes. To others, MEMS has become much broader, encompassing almost any fabrication technique performed on a small scale. In considering the future potential of MEMS, it is very common to compare the technology to the integrated circuit field, and to project growth based on the growth of the integrated circuit industry.[33] In making this comparison, however, it is important to consider some of the fundamentals that have led to the phenomenal growth of the integrated circuit field. We should realize that the ability of MEMS to mimic the performance, growth, and impact of the integrated circuit field will depend on the degree to which we understand and emulate the fundamental issues driving the growth of the integrated circuit field. The characteristics of integrated circuits that have led to their incredible impact are that they are cheap, they are powerful, and they are reliable. There are certain fundamentals in the integrated circuit field that have led to these three fundamental properties. It is well worth considering these fundamentals in considering strategies for the next silicon revolution. One of the first fundamentals of the integrated circuit industry is that there is a common technology base across many manufacturers, products, applications, and industries. The basic building block of virtually all integrated circuits is the MOS transistor. The function of the integrated circuit is determined by the way in which the transistors are connected. Over the course of the silicon revolution, there has been an unimaginable number of transistors made. With such an incredibly large number of transistors made by many different companies for many different products and applications, an enormous knowledge base has developed about how they work, how to make them, and why they fail. Because of the huge volume of this fundamental building block being made across the world, an extremely large science-based understanding of the operation of the transistor has developed from the atomic level on up. In addition to the science-based understanding, enormous empirical understanding exists for process recipes for fabrication as well as an empirical understanding of the do's and don'ts of manufacturing these devices. This synergy between competing manufactures also leads to large support for the advancement of the manufacturing tools. Companies that make integrated circuit manufacturing tools know that if they can create a more competitive tool, a huge worldwide manufacturing base will buy it. This helps to continue to drive the advancement of the quality of the manufacturing tool set. The fact that the basic transistor building block is the same for all competitors in the integrated circuit industry means that the companies that win are those that can: 1) make the transistors at the lowest cost with the greatest reliability, and 2) can connect them together through design to create products that accomplish functions useful to customers.

In addition to the common technology base, the integrated circuit industry relies on a common material set. Virtually all integrated circuits are made from the same very limited material set. Silicon, silicon dioxide, silicon nitride, aluminum, and tungsten are the materials in virtually all integrated circuits. Few successful products draw on any materials outside this limited set. Introduction of new materials such as ferroelectric materials,[34-35] other metals,[36-38] or other dielectric materials[37-39] have been tried on many occasions but with extremely limited success. Introduction of new materials is risky because it introduces new potential failure mechanisms such as material incompatibility problems, cross contamination problems, thermal incompatibility problems, and process integration problems.[40-43]

Another fundamental that has facilitated the success of the microelectronics industry has been the concept of batch fabrication. Integrated circuits are batch fabricated, thousands at a time, with no individual assembly or manipulation required. The cost of building a single high-performance transistor or 100 million transistors on a wafer is the same.[44] This concept of completely batch fabrication processes is one of the most compelling fundamentals that has resulted in the low cost, and hence high impact of microelectronic devices. It is critical to understand that the impact, of microelectronics comes from the fact that they are cheap, and the fact that they are inexpensive comes from the fact that they are batch fabricated. To the degree that any process sequence is introduced into the manufacturing flow that is not a batch process, the devices will no longer be cheap, and hence will have limited impact. These batch fabrication techniques also contribute to the high degree of process control achieved, which leads to high circuit yields, which also contributes to the low cost.

In now considering the future development of MEMS, it is important to contemplate some of the lessons of the integrated circuit field mentioned above. Today, MEMS does not draw on a common technology base. For the most part, MEMS technologies are application specific. For example, to make a display device, a company would be developing new processes, new technology, and possibly even new materials. To make an accelerometer, a company would develop different processes, different technology and different materials. For the future vision of the next silicon revolution to be fully realized, there is a need to develop industry standard MEMS technologies that can be used across different manufacturers and different products. Future product development should be centered around new designs in standard technologies, not new technologies for new applications. The realization of a common MEMS technology base will also enable manufacturing tool developers to begin to focus on MEMS specific manufacturing tools. Without a common and widely used MEMS technology base, few tool vendors will be interested in this field.

The early work in MEMS has also led to a proliferation of new materials, much broader than those commonly used in the integrated circuit world. The incorporation of new materials into MEMS technologies introduces a number of risks. First, one is no longer able to draw on the enormous data base that exists on the performance, process recipes, and failure mechanisms of traditional microelectronic materials. An additional risk is that of the mind set of the existing integrated circuit manufacturing infrastructure. Managers who are in charge of integrated circuit manufacturing facilities do not view non-traditional materials as new materials or new opportunities; they view them as contamination. Most existing MEMS development activities are working on the assumption that when they demonstrate relevant prototypes, full-scale production will be accomplished in some existing microelectronics fabrication facility.If the new MEMS technology or product includes non-standard materials, this greatly increases the risk that the technology will never be allowed inside the production facility.

It can be seen from the above discussion that there is a need for more standardization in MEMS technology. With the broad number and types of MEMS technologies, it is possible to speculate about which of these will become the standard. A taxonomy to consider the existing MEMS technologies can be built around two metrics. The first metric is how closely the technology mimics the fundamental characteristics of the IC industry, and the second metric is how large of a market (if any) is driving a particular technology. For example, a technology that does not posses many of the fundamental characteristics of integrated circuits, and is not driven by a high volume market, will probably not become the standard. On the other hand, a MEMS technology that contains all of the fundamental characteristics of integrated circuits, and is being driven by a large market, has a reasonable chance of becoming a standard. Since standards depend on the market pull, it is important to consider the near-term applications driving MEMS development. A number of market studies have been performed projecting large growth in the market for MEMS-based devices. A study performed by Systems Planning Corporation has projected that the market for MEMS devices will be $14 billion/year by the year 2000. It projects that these components will enable systems worth almost $100 billion/year. For the most part, such studies are done by looking at application areas that use sensors and actuators to calculate which of those application areas could be better served by MEMS-based devices and to then assume that MEMS would capture some share of that market. This is similar to the way projections were done in the early days of integrated circuits. When discrete transistors were first introduced as a replacement for vacuum tubes, people became very excited because they could see that radios that were built with tubes and were big and bulky could now be built much smaller, cheaper, and better using transistors. People projected markets of millions of dollars a year for transistors because of the possibility of small portable radios. These projections were correct in that they did correctly see the market for smaller cheaper radios. What the projections often failed to see, however, were the huge totally new markets enabled by the new technology, such as high-performance, affordable computers on virtually every desktop. In the same way, today's projections for the market for MEMS could very well be overlooking the totally new applications that are enabled by MEMS. Given that we understand the limitations of trying to predict the future, it is still interesting to consider some applications that seem certain for MEMS.

1.2 APPLICATION EXAMPLES

MEMS has progressed to the point that is clear that the technology will be more than simply a laboratory curiosity, and successful, high volume commercial products will be a reality. While MEMS is an emerging technology and has not yet come close to reaching its full commercial potential, there have been a number of high-volume products introduced recently that are MEMS based.

1.2.1 Displays

While a number of successful products and applications have been realized, it seems certain that display systems will be a major driver for the development of MEMS technologies. Today, information is transferred from electronic devices to people through display technology. The dominant display technologies today are CRTs and liquid crystal display systems. Low-cost, high-performance

MEMS-based display systems would have applications in television projection systems, computer displays, virtual reality headsets, pager displays, and practically any other product that could benefit from providing information to the user. This application area is particularly well suited for MEMS because it exploits some of the fundamental characteristics of MEMS. Like integrated circuits, it is possible to build large arrays of MEMS devices side by side, with each unit being identical.[45-48] This characteristic is ideal for displays, since an array of mirrors can be used to project an image, with each mirror representing one pixel of an image. The other factor that makes this application area attractive for MEMS has to do with the issue of contamination. Because MEMS are very small mechanical devices with minuscule features, they are extremely susceptible to particulate contamination. It is critical, therefore, for them to be protected from the small particulates that exist in normal air. For the display application, it is possible to package the MEMS device in something that amounts to a normal IC package with a transparent lid. The transparent lid can allow light in and out, but will protect the MEMS chip from particulate contamination.

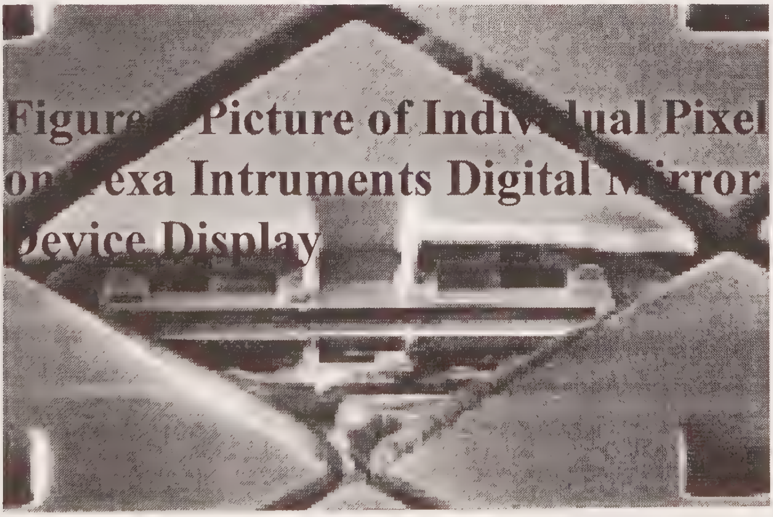

Figur Picture of Indiv ual Pixel
on exa Intruments Digital rror
evice Display

FIG. 1.2 Cross section of one pixel from Texas Instruments Digital Mirror Device.

MEMS are already making inroads into this important application area with Texas Instruments' introduction of its Digital Mirror Device (DMD).[46-47] The DMD (Fig. 1.2) is a projection system based on a very large array of micromachined mirrors. These mirrors are integrated with on-chip CMOS microelectronics which control the position and operation of the mirrors. The fabrication technology used to realize this product only requires standard integrated circuit materials, processes, and tools. Thus, it is completely compatible with standard integrated circuit infrastructure, helping to assure the manufacturability and performance of the product. A number of large screen projection systems currently on the market use this DMD chip as the heart of the system.

Other MEMS-based display systems are also under development. Work is underway at Silicon Light Machines on the development of a new type of MEMS-based display.[49] Instead of projecting images by tilting mirrors, this technology operates by the vertical displacement of small diffraction gratings. Each pixel of an image is represented by an independently controlled diffraction grating.

While there are a number of different approaches for using MEMS to create display systems, and each has its own set of advantages and disadvantages, it has become clear that MEMS-based displays will be a growing part of the market in the future.

1.2.2 Inertial Sensors

Another high volume market driving the development of MEMS is inertial sensors. Inertial sensors are sensors that measure acceleration and rotation. Accelerometers are used to measure acceleration, and gyros are used to measure rotation. Accelerometers combined with gyros provide the ability to measure arbitrary motion and position. Very large markets exist for chips that have an ability to measure motion and position. One of the earliest areas to incorporate MEMS devices is the application of airbag deployment. MEMS-based accelerometers are an ideal technology for this application.

Some of the requirements for this application are listed below. First, as you can imagine, reliability is very important. It is critical that in the event of an accident, the airbag is deployed. Similarly, it is just as important that the airbag not be inadvertently deployed when there is no accident. Such an event could prove to be very distracting to the driver. Since MEMS-based accelerometers are built with standard integrated circuit manufacturing techniques, it is anticipated that they should have the same high and predictable reliability levels commonly associated with integrated circuits. This makes them an ideal technology for this critical application area. Another critical factor for the airbag application is cost. In the highly competitive automotive industry, the cost of components is critical.[50-51] Most automotive sensors must be produced for well under $5.00 each.

The airbag deployment sensor represents one of the earliest uses of MEMS inertial sensors in automobiles. However, as the performance of the sensors increases, a number of new automotive applications will open up. For example, accelerometers can be used to measure vibration on a car,[52] and that information can be fed back to the active suspension, creating an active vibration control system to smooth the ride. Also, as MEMS products become available to measure rotation (gyros), it will be possible to make active skid suppression systems on cars. By measuring rotation or yaw, it is possible to tell that the driver is losing control of the car. With this information, the anti-lock brake system can apply the brakes to the correct individual tire, and bring the car back under control. Today, very expensive luxury cars have this feature.[53-56] However, the cost of the rotation sensors makes the technology impractical for mainstream cars. With the low cost expected from MEMS-based rotation sensors, it is anticipated that this important safety feature could become much more widely available.

There are many applications outside the automotive industry also driving the development of MEMS-based inertial sensors. There are many industrial applications that rely on large, rapidly spinning machines. In many of these applications, it would be beneficial to be able to more precisely measure the actual rotation of the machines. This information could be used for active machine control, or could be used to look for anomalous rotation signals, perhaps indicating a precursor to catastrophic failure. In many instances, if a machine could be turned off and repaired before catastrophic failure occurs, much money and time could be saved.

A number of commercial products will also be enhanced by low-cost, high-performance inertial sensors. It is anticipated that such devices would enable sensors to measure head and hand position for virtual reality input devices. The lack of affordable, reliable sensors to track head and hand position has limited the widespread use of virtual reality in home computers.

MEMS is well suited for use in creating a variety of inertial sensors.[57] Analog Devices has developed a successful commercial product,[58-59] the ADXL50, pictured in Fig. 1.3. This chip measures acceleration, and contains both a MEMS accelerometer, and the necessary CMOS microelectronics to control the accelerometer, creating, in effect, a complete integrated system on a chip. This chip is presently being used to deploy the airbag in a variety of new automobiles.

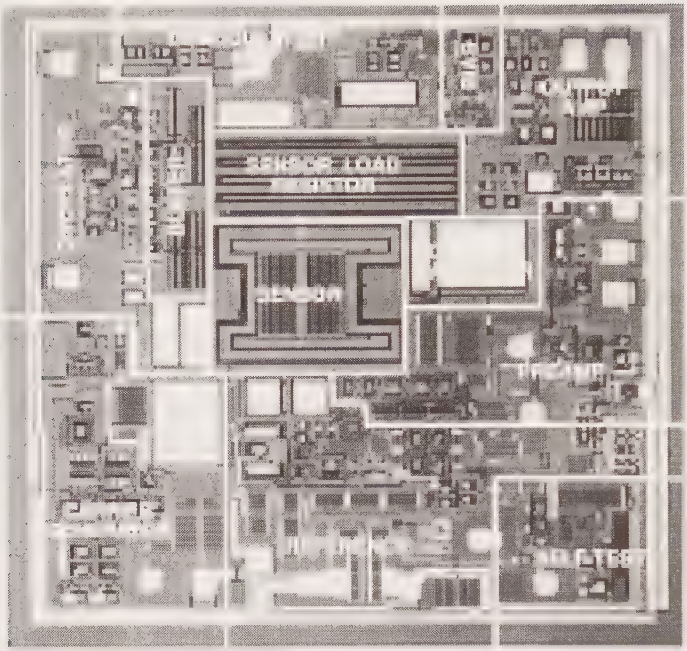


FIG. 1.3 Analog Devices' ADXL50 chip for measuring acceleration.

1.2.3 HP Ink Jet Print Head

One of the largest volume MEMS devices realized to date is the ink jet print heads.[60] This device is a MEMS-based component, and serves as the heart of many ink jet printers. These devices are comprised of an array of MEMS-based heater elements that are positioned behind simple orifices. When the heater is turned on, a bubble is formed in the ink, which shoots ink through the orifice.[107] As time has progressed, advances in the MEMS manufacturing technology has led to these components transitioning from black and white to full color, and with more precise ink drop sizes, the resolution continues to increase.[60]

1.2.4 Biomedical Instrumentation

One of the rapidly emerging applications fields for micromachining technologies is biomedical instrumentation. Micromachining technologies are of interest to the biomedical instrumentation industry because the manufacturing technologies offer an avenue to produce miniaturized instrumentation systems with sub-micron precision. Miniaturization of instrumentation is of interest for a variety of reasons, dependent on the application. For example, in the case of surgical instrumentation, micromachining technologies can be used to develop precision surgical tools for endoscopic and delicate surgical procedures. The need for precision surgical tools is driven by the need for less-invasive surgical procedures as well as by the number of specialized microscopic procedures under study and in practice today. Another example of miniaturized biomedical instrumentation is biological/chemical analysis systems. In the case of miniaturized analysis systems, scaling down of the system dimensions from "macro" to "micro" scale dimensions has several advantages. The key advantages of miniaturization are faster analysis times, highly parallel analysis systems, smaller sample/reagent volumes, and the opportunity to manufacture disposable analysis systems using batch fabrication technologies.

As is similar to other application areas of micromachining technologies, micromachining enables the realization of biomedical instrumentation previously unattainable using conventional manufacturing techniques. Therefore, micromachining technologies can be used to advance biomedical instrumentation development by enabling the development of novel instrumentation and by enabling miniaturization of existing biomedical instrumentation. Examples of novel biomedical instrumentation manufactured or under development using micromachining technologies include multichannel neural stimulation and recording systems, whole cell systems, and artificial sensory systems. Examples of miniaturized instrumentation developed as alternatives to larger"macro" scale instrumentation include biological/chemical analysis systems and medical monitoring systems.

The number of biomedical applications for micromachining technologies is rapidly growing. Since micromachining technologies are a relatively new set (and increasing set) of manufacturing technologies, there are many critical applications still to be addressed. To date, many advances have been made in biomedical instrumentation development as a result of using micromachining technologies to fabricate (or partially fabricate) the total instrumentation system. The main types of biomedical instrumentation that utilize micromachining technologies are neural stimulation/recording systems,[61-73] biological/chemical analysis systems,[74-89] whole cell systems,[90-95] surgical instrumentation,[96-98] medical monitoring systems,[99] phosthetic devices (e.g., artificial limbs),[100] and tactile sensors.[101-104] Each of these instrumentation types is briefly discussed below.

1.2.4.1 Neural stimulation/recording instrumentation

The development of biomedical instrumentation using micromachining technologies began in the mid- to late-1960s. Some of the first instrumentation developed was for neural stimulation/recording systems.[61,62] These neural stimulation/recording systems utilized silicon as the structural material and an integrated circuit approach to manufacturing the microelectrodes necessary for the neural interfacing. As can be imagined, there are a number of applications driving the development of neural interfaces. Some of the applications are: (1) general mapping of the nervous system, particularly the central nervous system regions (e.g., the brain), to understand the function of (and subfunctions within) regions of neurons; (2) "hands-free" prosthetic limbs controlled by thought processes in the brain; (3) neural interfaces for artificial sensory systems; (4) partial reactivation of paralyzed limbs; (5) limited control over organ and gland functions; and (6) auditory aids for the deaf.[101]

Over the past several decades, scientists and engineers have devised a variety of innovative techniques to make microelectrodes that can detect extracellular neural signals or stimulate neurons at specific sites. Some of the earliest work in the area of extracellular recording of neurons involved the use of individual metallic microelectrodes fabricated using small-diameter wire.[105-106] The tips of these electrodes were formed using electrochemical sharpening techniques and insulated by immersing the electrodes in a lacquer-based solution. The electrode recording sites were established at the tip by removing the insulative layer using a high-voltage discharge. Since the area of the recording site is a primary factor in determining the electrical characteristics of metallic microelectrodes, this technique often led to significant differences in the characteristic impedance from one electrode to the next due to variations in the recording site area.

The first microelectrodes fabricated using micromachining techniques were constructed of gold electrodes supported on thin silicon support beams.[61] The length of the electrode shafts were in the range of 10 μm to 50 μm with silicon dioxide serving as the electrode insulating material. Later, this process was revised to include multichannel probes fabricated on silicon shanks with the integration of signal amplification and multiplexing circuitry.[62-66] Similar work has been performed by others, including multielectrodes that use polyimide as the electrode insulating material and use plasma etching to define vias in the polyimide for recording site areas.[67]

Other microelectrodes fabricated using micromachining techniques include the diving-board electrode[68] for chronic in-vitro recording and stimulation, electrode arrays fabricated on metallic substrates,[69] the cone electrode,[70] the metallic microelectrode arrays (see Fig. 1.4),[71] and various others for chronic applications[72,73] in which nerve regeneration is required.

1.2.4.2 Biological/chemical analysis systems

One of the largest biomedical application areas for micromachining technologies is biological/chemical analysis systems. Over the past ten years, micromachining technology has been used to develop a large number of miniaturized analysis systems. Typically these microanalysis systems include microchannels and microchambers as an essential part of the design in order to control sample flow, mixing, and analysis operations. Currently, micromachined fluidic systems include gas chromatography,[74-76] liquid chromatography,[77,78] free-flow fractionation,[79,80] electrophoresis,[81-84] and electrical field flow fractionation.[85] Other analysis techniques, such as polymerase chain reaction (PCR), have recently been miniaturized as well.[86-89] The primary advantages of miniaturizing analysis systems are increased resolution of the analysis results, reduced analysis times, smaller sample and reagent volumes, the opportunity to manufacture a disposable analysis system, and increased manufacturing precision. The primary hurdles being faced by industry interested in developing products based on the miniaturized analysis system approach are system integration and interfacing with macro-scale sample preparation systems and detection systems. Currently, there are a number of successful products being marketed that utilize micromachining technologies, including blood analysis systems, glucose monitoring systems for diabetics, and polymerase chain reaction systems.

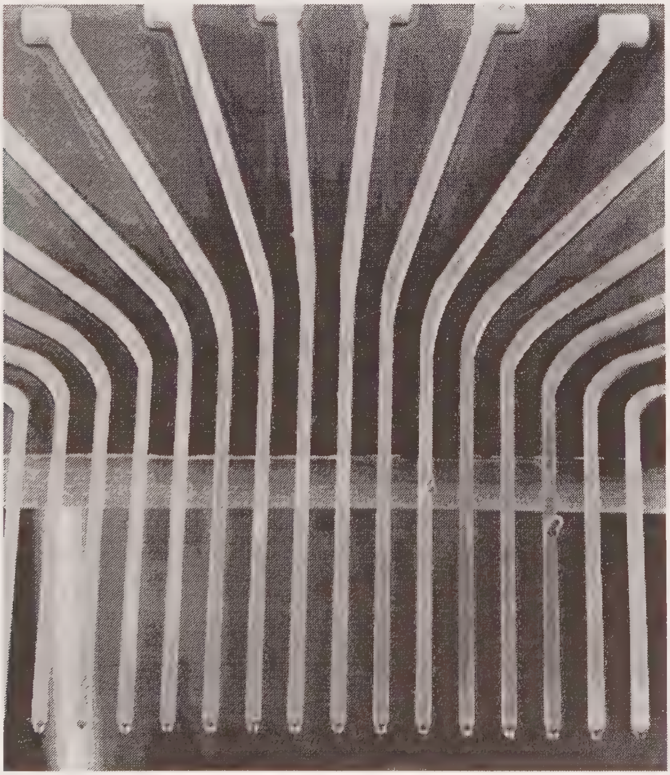


FIG. 1.4 A metallic microelectrode array. The array has 16 isolated electrodes encapsulated in silicon nitride except for the recording tips and bonding pads. The electrodes are 15 μm thick, 20 μm wide, and the center to center spacing is 75 μm.

1.2.4.3 Whole cell systems

Another application area for micromachining technologies is in the development of whole cell systems. Applications of whole cell systems generally fall into one of the four following categories: cell-sorting systems,[90-92] active cell characterization systems,[92] cell growth systems,[93] and microsystems for cell assays.[94-95] Cell-sorting systems represent a new application area for micromachining technologies. The goals of the present work are to develop systems capable of sorting cells due to variations in the physical and electrochemical nature. Various prototype systems have been fabricated and undergone initial systems testing, but final cell-sorting systems capable of massively parallel operation are still under development. The applications of cell-sorting systems include cell-counting systems for determining red/white blood cell counts, cellular screening tests for disease and cancer detection, and cell aging/storage evaluation systems for monitoring storage of biomaterials (e.g., blood in hospitals and blood banks).

Active cellular characterization systems is another new application field for micromachining technologies (see Fig. 1.5).[92] Active cellular characterization systems are being developed to study the cellular processes within individual cells such as the variation on cellular impedances due to environment changes. In addition, active cellular characterization systems are being developed for controlling membrane potentials and for cellular imaging. Cellular assay systems are another interesting application field of micromachining technology. Miniaturized cell assay systems can be used to perform assays on individual cells or small groups of cells using a precisely controlled environment. The miniaturized systems for microscale assays can be a complex system capable of multiple steps in which cells are guided through various microchambers for mixing with reagents and other biochemicals such as inhibitors and activators.

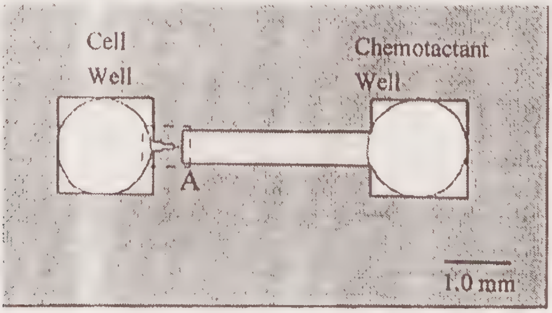


FIG. 1.5 A schematic of the whole cell analysis system. The system includes a glass cover slip and a polymer channel defined on a glass substrate. The area A contains a 4 × 6 × 150 μm^3 microchannel.

1.2.4.4 Precison surgical tools and medical monitoring instrumentation

Two related classes of biomedical instrumentation utilizing micromachining technologies are precision surgical tools and medical monitoring instrumentation. Precision cutting tools are being developed both for commercial applications and for use in the research laboratory.[96-98] One commercial application of precision cutting tools is endoscopic instrumentation for endoscopic surgical procedures. In this case, as in the case of all endoscopic procedures, the patient undergoes surgery without the need for large skin incisions. The surgical procedures are performed using miniaturized tools inserted into the body region of interest through a sheath that extends from the exterior to the interior. Real-time visual monitoring of the procedures is accomplished using a fiber optic feed into the body. There are many cases in which it would be desirable to transect tissue with a precision mechanical cutting instrument capable of operating in biological fluids and within tubes and vessels. In addition, the precision cutting instruments must have simple operation, minimize damage to surrounding tissue, and allow for a controlled depth of the incision. An important example of an application for the precision mechanical cutting instrument is the removal of extraneous tissue within the Fallopian tubes of females. Obstructed Fallopian tubes are a major cause of infertility in females. Typically, die tests are performed to diagnose the problem, but there is currently no procedure available to correct the problem. In order for the females to conceive, the individual must rely on artificial insemination. Another example of a far-reaching application for the precision cutting instrument is removal of plaque from obstructed arteries. In this application, the cutting instrumentation is used in combination with other endoscopic instrumentation to remove obstructions inhibiting blood flow through the cardiovascular system. In addition to using the precision microcutting tools for surgical procedures, there are applications for such tools in the research laboratory. These tools can be used for preparing small samples of plant and animal tissues for analysis, to study mechanical properties (such as shear strength) of various tissue and cartilage, and for detailed layer-by-layer analysis of selected tissues.

Medical-monitoring instrumentation is another related application area for micromachining technologies. Some examples of medical-monitoring instrumentation for which micromachining is utilized include active catheter systems,[99] external EKG monitoring systems,[99] and biological/chemical analysis systems.[74-89] In the case of active catheter systems, micromachining technology is used to manufacture numerous sensory components including pressure sensors, temperature sensors, and cardiac output monitors. Pressure sensors are used to monitor the blood pressure internally at specific locations (e.g., inside the pulmonary artery). Temperature sensors are used to monitor the blood temperature (e.g., inside the heart). Flow rate detectors are used to measure the cardiac output by measuring the velocity of the blood through a specific blood vessel or artery. In the case of external EKG monitoring systems, micromachined pressure sensors are used to measure the EKG by externally positioned sensors on the body. Biological/chemical analysis systems were discussed in a previous section.

1.2.4.5 Prosthetic devices and tactile sensors

Micromachining has applications in the development of advanced prosthetic devices and tactile sensors. In the case of prosthetic devices, micromachining technology can be used produce sensors and actuators. The sensors are necessary for obtaining input from the surrounding environment for the prosthetic devices. One example is cochlear prosthesis.[100] In this case, micromachining technologies are used to fabricate an implantable neural interface for the auditory system. Another example is tactile sensors.[101-104] Tactile sensors can be used for a variety of applications ranging from the development of advanced prosthetic limbs to artificial sensory systems capable of providing environmental input (e.g., textures) to a human via the sense of touch. Another example of utilizing micromachining technologies to produce tactile sensors is for developing systems for reading embossed characters (e.g., credit cards, debit cards, card accessing systems).

1.3 FABRICATION TECHNOLOGIES

From the above examples, it can be seen that a variety of different products can be successfully realized using MEMS fabrication technologies. While there are as many different MEMS fabrication technologies as there are MEMS products on the market, these technologies can be boiled down to three basic techniques. In the following sections, we discuss these basic fabrication technologies.

1.3.1 Bulk Micromachining

Bulk micromachining is one of the earliest forms of micromachining, and is suitable for creation of a variety of sensor elements. While there are many different types of bulk micromachining, the basic concept is to create mechanical features by removing material from a wafer.[44] The basic silicon wafer serves as the mechanical material, and the structures are created by removing the parts of the wafer that are not wanted. Typically this is done through a sequence of photopatterning steps followed by wet etching.[108-113] The wet etching removes the unwanted part of the wafer, leaving the desired mechanical features behind. Such techniques have been used to realize a variety of successful commercial products, including pressure sensors[114] and accelerometers.[115] The advantage of bulk micromachining is that it can be done with a minimum of capital equipment, and is well suited to the fabrication infrastructure found in a number of universities.

1.3.2 Surface Micromachining

Surface micromachining is a technology that enables the fabrication of a variety of actuators[116] or motors,[117] in addition to a number of physical[1-9] and chemical sensors.[10-14] Surface micromachining is more complex than bulk micromachining, and requires more fabrication infrastructure. While it requires more fabrication infrastructure, it enables the realization of much more complex structures than what is possible in bulk micromachining. In surface micromachining, mechanical devices are formed by the successive deposition, photopatterning, and etching of thin films on the surface of the wafer.[44] Two types of thin films are used: structural films and sacrificial films. The sacrificial films serve as spacer layers between the structural levels. As a final step, the sacrificial layers are etched away, leaving the structural levels free to move and operate. Typically in surface micromachining, polycrystalline silicon is used as the structural material, and silicon dioxide is used as the sacrificial material. At the end of the process, wafers are immersed in hydrofluoric acid, which quickly removes the silicon dioxide layers, leaving the structural polycrystalline silicon untouched.

It is possible to use other structural/sacrificial material combinations in surface micromachining. Some products, including the Texas Instruments Digital Mirror Device, are created using aluminum as the structural material, and photoresist as the sacrificial material.[47]

1.3.3 Mold Micromachining

A third fabrication technology for the creation of MEMS devices is mold micromachining. In this technology, mechanical features are created by first creating a mold. This mold is in the shape of the desired mechanical structure. The mold is then filled with a structural material to create the desired mechanical devices. The mold can then be etched away, or the mechanical device can be removed from the mold.

The most famous technology for mold micromachining is the LIGA technology.[118] LIGA is an acronym for the German words: lithography, electroforming, and molding. LIGA creates very small, very precise mechanical structures in the following way. Masks are created containing the shape of the desired mechanical structures. Then, a synchrotron radiation source is used to expose a thick resist layer. The irradiated area of the resist layer is then dissolved by use of an appropriate developer, leaving a mold in the shape of the desired mechanical structure. Electroplating of Cu, Ni, or other metal into the mold then creates the desired mechanical structure. Following this, the remainder of the resist is dissolved, leaving the mechanical structure.

While the LIGA technology enables the fabrication of high-aspect-ratio micromachines, and creates extremely precisely dimensioned devices, the technology is limited in that it requires a high-energy synchrotron radiation source. Such a source is not part of the standard integrated circuit infrastructure, making this technology not widely available outside of a few research organizations.

Recently, many other mold technologies have been developed that are much more compatible with the traditional integrated circuit infrastructure. The HEXSIL technology[119-120] developed by the University of California at Berkeley creates a mold in a silicon wafer by deep reactive ion etching. The mold is then lined with a sacrificial material, and then filled bychemical vapor deposition of a suitable mechanical material, such a polycrystalline silicon. The lining layer of the mold is then chemically dissolved, leaving behind the desired mechanical structure.

1.4 MOTIVATION TO INTEGRATE MEMS AND MICROELECTRONICS

Using the MEMS manufacturing technologies above, or variations of these, it is possible to create a variety of microscopic sensors and actuators. However, simply using these techniques to fabricate components is to not realize the full potential of MEMS. If we look at many sensor systems, the cost of the system is often dominated by the cost associated with packaging and assembling the individual components. A basic MEMS sensor component might only cost a few dollars, but the cost of a useful sensor system incorporating that sensor component might be hundreds or thousands of dollars. The true power and impact comes when we begin to integrate the MEMS and CMOS control electronics on the same piece of silicon, enabling the batch fabrication of entire systems on a chip. Since it is possible to batch fabricate the entire system on a single chip, the costs associated with packaging and assembly are greatly reduced, leading to extremely powerful, low cost solutions for many sensing applications.[121]

While the integration of MEMS and sensors can lead to inexpensive systems on a chip, there are a number of factors in addition to cost driving this trend toward integrated devices. In many cases, system performance can be significantly increased by integration of MEMS and microelectronics on the same piece of silicon.[121] Many MEMS-based sensors have capacitance as the output parameter.

For a nonintegrated system where the control electronics are on one chip and the MEMS-based sensor on another, the sensitivity and performance can be affected by noise and parasitics. For example, a nonintegrated accelerometer requires large bond pads to enable electrical connection to the outside world. In many cases, the parasitic capacitance of this bond pad can be several orders of magnitude greater than the output signal from the accelerometer itself. This creates a voltage divider where most of the sensor signal is dropped across the parasitic capacitance associated with the bond pad. These parasitic capacitances can have a very adverse effect on sensitivity and performance of the system. The integration of electronics and MEMS on the same piece of silicon almost eliminates this interconnect capacitance, enabling much higher sensitivity and performance.

There are additional factors driving the integration of MEMS and microelectronics. One of the primary failure mechanisms in traditional integrated circuits is failure of bond wires. The more bond wires in a given product, the greater the risk of bond wire failure, and hence the lower the system reliability. By integrating MEMS and microelectronics on the same piece of silicon, the number of bond wires are minimized, reducing the probability of failures due to bond wire problems.[122]

While it can be seen that there are compelling reasons to integrate MEMS and microelectronics, it is by no means a trivial or straightforward proposition. There are many challenges associated with the development of integrated technologies. Problems with material incompatibility, process incompatibilities, and thermal budget incompatibilities create significant challenges in creating high-performance integrated technologies.[123]

1.4.1 Integration Strategies

As an example of the challenges encountered in the development of integrated MEMS/CMOS technologies, let us consider the case of the integration of traditional polycrystalline silicon surface micromachines with CMOS control electronics. The need for high-performance, low-cost inertial sensor systems on a chip have been a driving application for this integration effort. While it is clear that there is a strong application drive to performing this integration, there are many different integration strategies that can be pursued, each with its own advantages and disadvantages. The first issue that must be addressed is the fabrication sequence, that is, the order in which the micromachines and microelectronics will be fabricated. Below we examine some of the integration strategies that have been tried to date.

1.4.2 Microelectronics First, Micromachines Last

One of the first technologies that integrated surface micromachining and microelectronics on the same piece of silicon was developed[124] at the University of California at Berkeley at the Berkeley Sensor and Actuator Center. The integration strategy for this technology was to build the microelectronics first, followed by the surface micromachined devices (Fig. 1.6). The motivation for this strategy was that it enabled a modular approach to integration. One could envision building the microelectronics, possibly at a foundry, followed by the fabrication of the micromachines, possibly at a university or research facility. This technology was successfully developed at UC Berkeley, and it demonstrated the motivation for, and feasibility of, integration of MEMS and microelectronics.

This technology, however, had several issues that made its incorporation into a large-scale manufacturing environment a challenge. The integration challenges with this strategy were centered around the thermal budget needed to fabricate the micromachines. In order to achieve flat MEMS devices, the structural material must be virtually stress free. Any residual stress in the structural material will cause the devices to curl up when the final release process is performed. In polycrystalline silicon surface micromachining, high-temperature anneals are used to eliminate the stress in the film.[125] In some cases, temperatures as high as 1000°C or more are desirable. The problem is that these high temperatures, necessary to relieve stress in the micromachined devices, damage the underlying CMOS microelectronics. One of the major problems is that the traditional interconnect material used in CMOS is aluminum. Aluminum melts at about 450°C. In order to get around this problem, the research team used tungsten as a replacement for aluminum, because of tungsten's relatively high melting temperature. Even with the tungsten interconnect, the high-temperature anneals caused very high contact resistance in the source/drain contact regions.

The high temperatures can also create a problem by causing the source/drain regions to diffuse too far under the gate region of the CMOS transistors. Attempts to lower the temperature of the anneal of the polycrystalline silicon were somewhat successful, but in the end it was not possible to build high-performance CMOS with the best possible MEMS structures because of these thermal budget problems.

1.4.3 Micromachines in the Middle

In order to overcome some of the manufacturing challenges associated with the "MEMS Last" integration strategy developed by the University of California at Berkeley, Analog Devices Incorporated (ADI) has developed an integration strategy that involves building the micromachines in the middle of the process.

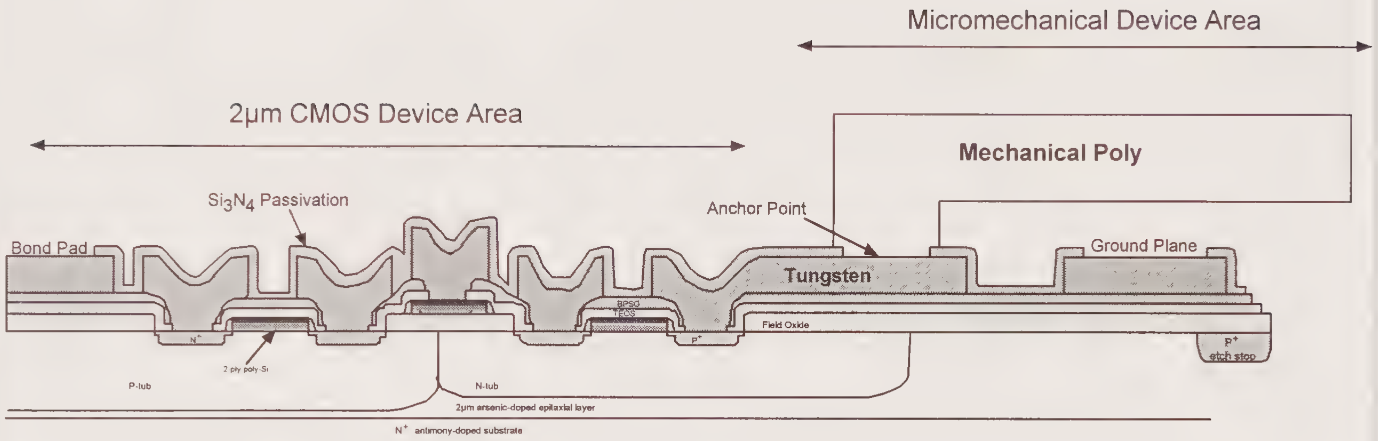

Micromechanical Device Area
2µm CMOS Device Area
Mechanical Poly
Si_3N_4 Passivation
Anchor Point
Bond Pad
Ground Plane
Tungsten
BPSG
TEOS
Field Oxide
N^+
P^+
P^+
etch stop
2 ply poly-Si
P-tub
N-tub
2µm arsenic-doped epitaxial layer
N^+ antimony-doped substrate

FIG. 1.6 Cross section of the "MEMS Last" integration strategy developed by the University of California at Berkeley.

This strategy interleaves the micromachine and microelectronic fabrication steps in order to improve the manufacturability, and to overcome some of the problems associated with the thermal budget necessary to eliminate stress in the micromachines. The motivation driving this development activity was primarily the airbag deployment application. In this application, integration was critical from the perspective of price, performance and reliability.

This technology achieved integrated devices by interleaving the process steps of the microelectronics and micromachines. The process starts by performing the microelectronic processing steps. Microelectronic processing continues up to the point of deposition of the metal interconnect layer. Prior to fabrication of the interconnect layer, the micromachines are built. Following the fabrication of the micromachines, microelectronic fabrication steps resume with the completion of the metal interconnect and the passivation layers. By forming the micromachines prior to the metal interconnect, the high-temperature stress reduction anneals are done before the metal is put down, eliminating the need to abandon aluminum for a nonstandard interconnect material. An additional manufacturing advantage is that the contacts between the metal interconnect and the source drain regions of the transistors occur after the high-temperature anneals, resulting in excellent contact resistance parametrics. This "Interleaved" integration strategy overcomes some of the major manufacturing issues of the "Micromachines Last" integration strategy. ADI has put their "Interleaved" integrated technology into full-scale production, and are producing high volumes of intelligent accelerometer "systems on a chip" for automotive and other applications.

While this "Interleaved" technology overcomes the major obstacles in creating a manufacturable technology, it does not necessarily represent the ideal solution to the problem. While it has been demonstrated that the technology is suitable for large-scale manufacturing, several challenges still exist. The first is that while the high-temperature anneals are done prior to metalization, they are done after the transistors' source/drain regions are formed. The implanted dopants in these source/drain regions diffuse out in response to the high-temperature anneals. This means that this integration strategy is only suitable for somewhat large transistor dimensions. For modern high-density, submicron CMOS technologies, the high-temperature anneals would result in the source/drain regions diffusing together underneath the gates of the transistors, leading to circuit failure. Another challenge with such an integration strategy is that the technology is no longer modular. Now the processing for the micromachines and microelectronics are intertwined together and interdependent. Change in the processing for one would likely affect the other. This means that advancement to a newer microelectronic technology could require a large amount of re-engineering of the technology.

1.4.4 Micromachines First

In looking at the pros and cons of micromachines last and micromachines in the middle, the next option to consider is the option of building the micromachines first. The motivation for building the micromachines first is that if they are built prior to any microelectronics processes, one would be able to do any thermal anneal without a fear of affecting the performance of the CMOS. The obvious strategy of simply building the micromachines followed by building the microelectronics does not work well. The problem is that the integrated circuit manufacturing tool set expects to see perfectly flat wafers. Significant topography on the wafers creates untenable manufacturing problems. Simply fabricating micromachines on the surface of the wafer generates so much topography that large problems develop in the remainder of the fabrication sequence. The topography generates problems in achieving suitable photoresist coverage. Other problems include the formation of residual, parasitic strings of material after etching known as “stringers”. These problems had prevented any real progress in achieving a “Micromachine First” integrated technology.

Progress has been made recently at Sandia National Laboratories in a “Micromachine First” integration strategy.[126-127] (See Fig. 1.7.) This technology begins by etching trenches in the surface of the silicon wafer. The micromachines are then built in the bottom of these shallow trenches. The trenches are backfilled with a sacrificial silicon dioxide layer, and the wafer is then planarized completely flat. At this point, there are no microelectronic structures, so one is free to perform any thermal anneal sequence of interest, enabling the independent optimization of the stress properties of the micromachines. After performing the desired high-temperature anneal, we still have a flat wafer, so to the microelectronics fabrication tools, the wafer looks like any other wafer. Since the wafer is flat, it is possible to incorporate any modern, high-performance microelectronics technology of interest onto the wafer.

This micromachine-in-a-trench integration strategy allows the independent optimization of the micromachines and microelectronics. It is truly modular in that advancing to the next generation of high-performance microelectronics technology will not affect the embedded micromachined devices. Similarly, changes in the embedded micromachines do not affect the microelectronics. This strategy overcomes the final manufacturing hurdles to enable the seamless integration of high-performance micromachines and high-performance microelectronics on the same piece of silicon.

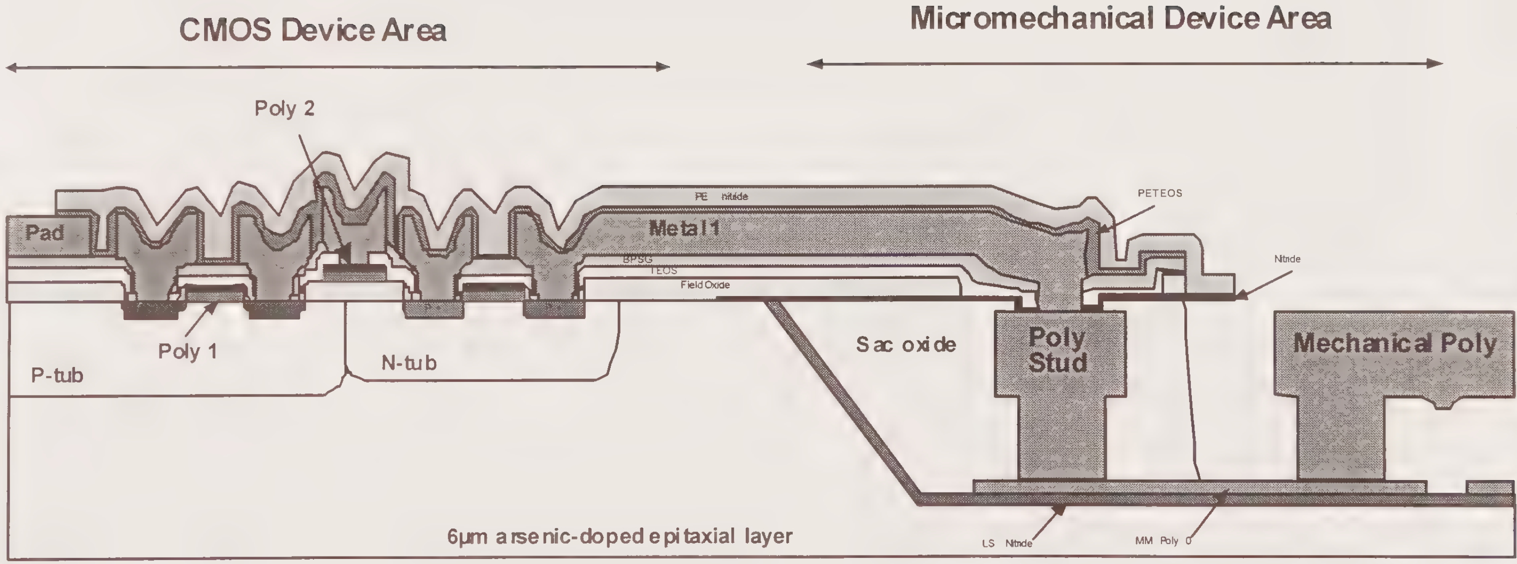

CMOS Device Area
Micromechanical Device Area
Poly 2
Pad
PE nitride
Metal 1
PETEOS
BPSG
TEOS
Field Oxide
Nitride
Poly 1
P-tub
N-tub
Sac oxide
Poly Stud
Mechanical Poly
6µm arsenic-doped epitaxial layer
LS Nitride
MM Poly 0

FIG. 1.7 Cross section of Sandia National Laboratories "Micromachine First" integrated technology.

1.4.5 Considerations in Determining Integration Strategies

As can be seen in the above case study of the integration of surface micromachining and microelectronics, there are many trade-offs to be considered in developing an integration strategy. It is sometimes difficult to determine up-front the correct strategy. Even with the most careful analysis, there can be unforeseen difficulties. It can be seen that significant resources were expended over a number of years by the University of California at Berkeley, Analog Devices Incorporated, and Sandia National Laboratories to investigate and understand the pros and cons of these three different integration strategies. We can see that the issues of MEMS topography were pushing the MEMS toward the end of the process, while the issues of thermal anneals were pushing the MEMS toward the front of the process.

It is interesting to consider that in each of these cases, only standard microelectronic materials were used. All of the trade-offs and challenges were centered around difficulties associated with topography and thermal budgets. Integration can become even more difficult when one begins to deal with nonstandard materials. In the case of nonstandard materials, not only are the above issues a concern, but now a number of new issues are introduced associated with contamination. Microelectronics operate based on the careful control of the conductivity of the silicon material. Dopants are introduced that precisely define the conductivity of the material. Any contamination introduced into the process in even minuscule amounts can change the desired conductivity and electrical properties of the CMOS transistors, resulting in nonfunctional devices. This makes it very difficult to conduct experiments with new materials in most microelectronic fabrication facilities. The fear is that investigation of new materials would not only destroy the microelectronics on the wafer of interest to the study, but could contaminate the processing tools themselves, killing the microelectronics on other traditional products sharing these same tools. The fear of yield problems due to possibly even undetectable amounts of contamination on the fabrication tools makes it virtually impossible to introduce any nonstandard materials into a fabrication facility responsible for delivering real microelectronic products. This contamination fear should be understood up-front when doing research into any new MEMS technology or integration strategy. If new materials are to be incorporated, it is best to be done at the end of the process. Then, there is the possibility of fabricating the microelectronics in a traditional integrated circuit fabrication facility, followed by the incorporation of the structures made of nonstandard materials at a university or research facility.

With the sometimes seemingly untenable problems associated with the creation of these integrated technologies, a temptation can exist to simply build discrete MEMS devices. One could build MEMS on one chip, microelectronics on another, and then wire bond them together. In some low-volume applications, this might indeed be the best strategy. However, for high-volume, high-performance applications, one can anticipate that integrated technologies will always displace hybrid technologies. If we look at the traditional microelectronics market, single-chip solutions to problems virtually always displace multiple-chip solutions to problems. If we look at the development of any number of consumer electronic devices from desktop computers to cellular phones, the undeniable trend over the last three decades has been toward reduction in the number of chips needed to perform a given function. It is anticipated that this trend will also exist in the world of MEMS. It is interesting to note that both the Texas Instruments DMD display devices and the Analog Devices ADXL50 are built in integrated technologies, with the micromachines and the microelectronics on the same piece of silicon. However, because of the large investment needed to develop integrated technologies, they are only suitable when the market for the product is large enough to recoup the start-up cost. It is anticipated that in low-volume market areas, hybrid devices represent the most suitable approach.

1.5 FUTURE POTENTIAL

These emerging applications and technologies show the potential of the MEMS technology. They represent only the tip of the iceberg, but are nonetheless important demonstrations. These successes show that MEMS can make the transition from laboratory demonstrations to viable commercial products. It is anticipated that such successful MEMS-based products will be introduced into the market at an ever-increasing rate. As this happens, MEMS will become ubiquitous. They will become part of many consumer products. They will make existing products better, and eventually will enable totally new products not even imagined today. The ability to give integrated circuits the capability to understand where they are and what is going on around them will likely impact the next century as profoundly as the integrated circuit's ability to think has impacted the present century.

REFERENCES

1. A. Tanaka, S. Matsumoto, N. Tsukamoto, S. Itoh, T. Endoh, A. Nakazato, Y. Kumazawa, M. Hijikawa, H. Gotoh, T. Tanaka, N. Teranishi, “Silicon IC process compatible bolometer infrared focal plane array,” *8th International Conference on Solid-State Sensors and Actuators and Eurosensors IX. Digest of Technical Papers*, **2** (1995), pp. 632–635.
2. C. H. Mastrangelo and R. S. Muller, “A constant temperature gas flowmeter with a silicon micromachined package,” *Tech. Digest 1988 Sol.–State Sensor and Actuator Workshop*, (1988), pp. 43–47.

3. H. Guckel, "Surface micromachined pressure transducers," *Sensors and Actuators A*, **28** (1991), 133–146.
4. J. J. Sniegowski, "Design and fabrication of the polysilicon resonating beam force transducer, " *Ph.D. Dissertation, The University of Wisconsin-Madison,* (1991).
5. W. Yun, R. Howe and P. Gray, "Surface micromachined, digitally force-balanced accelerometer with integrated CMOS detection circuitry," *Proc. of the IEEE Solid-State Sensor And Actuator Workshop '92*, (1992), 126.
6. W. Clark, R. Howe, and R. Horowitz, "Z-axis vibratory rate gyroscope," *Micromachining Workshop III: Technology and Applications*, (1996).
7. E. H. Klaassen, G. T. A. Kovaks, "Integrated thermal conductivity vacuum sensor," *Solid-State Sensor and Actuator Workshop*, (1996), pp. 249–252.
8. T. Stowe, K. Yasumura, T. Kenny, D. Botkin, K. Wago, and D. Rugar, "Ultrasensitive vertical force probe for magnetic resonance force microscopy," *Tech. Digest 1996 Sol.-State Sensor and Actuator Workshop*, (1996), pp. 225–230.
9. Gupta, d. P. Neikirk, "Design of an inductive proximity sensor using a two-coil planar transformer," *Proc. SPIE* **2642**, (1995), pp. 173–182.
10. Manginell, J. H. Smith, A. J. Ricco, D. J. Moreno, R. C. Hughes, R. J. Huber, S. D. Senturia, "Selective, pulsed CVD of platinum on microfilament gas sensors," *Tech. Digest 1996 Sol.-State Sensor and Actuator Workshop*, (1996), pp. 53–56.
11. Krebs and A. Grisel, "A low power integrated catalytic gas sensor," *Sensors and Actuators B*, *13-14* (1993), pp. 155–158.
12. Semancik, R. E. Cavicchi, K.G. Kreider, J. S. Suehle, P. Chaparala, "Selected-area deposition of multiple active films for conductometric microsensor arrays," *8th International Conference on Solid-State Sensors and Actuators and Eurosensors IX. Digest of Technical Papers*, **1** (1995), pp. 831–834.
13. Zanini, J. H. Visser, L. Rimai, R. E. Soltis, A. Kovalchuk, D. W. Hoffmann, E. M. Logothetis, U. Bonne, L. Brewer, O. W. Byrnum, M. A. Richard, "Fabrication and properties of a Si-based high sensitivity microcalorimetric gas sensor," *Tech. Digest 1994 Sol.-State Sensor and Actuator Workshop*, (1994), pp 176–179.
14. G. Hughes, "Chemical gas sensors on silicon," *Proc. SPIE* **2642**, (1995), pp. 104–109.
15. H. Comtois, V. M. Bright and M. W. Phipps, "Thermal microactuators for surface-micromachining processes," *Proc. SPIE* **2642**, (1995), pp. 10–21.
16. P. Lee, D. R. Ciarlo, P. A. Krulevitch, S. Lehew, J. Trevino, and M. A. Northrup, "A practical microgripper by fine alignment, eutectic bonding and SMA actuation," *8th International Conference on Solid-State Sensors and Actuators and Eurosensors IX. Digest of Technical Papers*, (1995), pp. 368–371.

17. J. Sniegowski and E. Garcia, "Surface-micromachined geartrains driven by an on-chip electrostatic microengine," *IEEE Electron Device Letters*, **17**, (1996), 366.
18. Fan, Y. Tai, and R. S. Muller, "IC-processed electrostatic micromotors," *Sensors and Actuators*, **20** (1989), pp. 41–47.
19. C. Tang, T. -C. H. Nguyen, and R. T. Howe, "Laterally driven polysilicon resonant microstructures," *1989 International Workshop on Micro Electromechanical Systems (MEMS 89),* (1989), 53.
20. J. J. Kruglick, S. Damle, K. S. J. Pister, "Three-dimensional structures for micro-optical mechanical systems in standard CMOS," *Proc. SPIE* **2642**, (1995), pp. 33–39.
21. A. Miller, G. W. Burr, Y. -C. Tai, D. Psaltis, S. -M. Ho, and R. R. Katti, "Electromagnetic MEMS scanning mirrors for holgraphic data storage," *Tech. Digest 1996 Sol.-State Sensor and Actuator Workshop*, (1996), pp. 183–186.
22. P. Deimel, Micromachined devices for optical applications, *8th International Conference on Solid-State Sensors and Actuators and Eurosensors IX. Digest of Technical Papers*, (1995), pp. 340–343.
23. T. Frank, "Development and manufacturing of a two-dimensional microactuator for moving an optical fiber," *Proc. SPIE* **2882**, (1996), pp. 226–235.
24. L. Huang, H. Zhang, E. S. Kim, S. G. Kim and Y. B. Jeon, "Piezoelectrically actuated microcantilever for actuated mirror array application," *Tech. Digest 1996 Sol.-State Sensor and Actuator Workshop*, (1996), pp191–195.
25. Smith, R. Carson, C. Sullivan, and G. McClellan, "Monolithic integration of waveguide structures with surface-micromachined polysilicon actuators," *SPIE Proc.*, **2722**, (1996), pp. 75–81.
26. N. Carr, S. Chamarti, and X. Gu, "Integrated pressure sensor with remote power source and remote readout," *8th International Conference on Solid-State Sensors and Actuators and Eurosensors IX. Digest of Technical Papers*, (1995), pp. 624–627.
27. Nardin and K. Najafi, "A multichannel neuromuscular microstimulator with bi-directional telemetry," *8th International Conference on Solid-State Sensors and Actuators and Eurosensors IX. Digest of Technical Papers*, (1995), pp. 59–62
28. R. Manginell, J. H. Smith and A. J. Ricco, "An overview of micromachined platforms for thermal sensing and gas detection", *Proc. SPIE* **3046**, (1997), pp. 273–284.
29. E. Cavicchi, J. S. Suehle, P. Chaparala, K. G. Kreider, M. Gaitan, and S. Semancik, "Micro-hotplate gas sensor", *Tech. Digest 1994 Sol.-State Sensor and Actuator Workshop*, (1994), pp. 53–56.

30. Smith, S. Montague, J. Sniegowski, J. Murray and P. McWhorter, "Embedded micromechanical devices for the monolithic integration of MEMS with CMOS," *Proc. IEDM '95*, (1995), pp. 609–612.
31. K. Ghandhi, *VLSI Fabrication Principles, 2nd Ed.* John Wiley and Sons, (1994).
32. R. Runyan, *Semiconductor Integrated Circuit Processing Technology*, Addison-Wesley, (1990).
33. See, for example, studies by SEMI, SPC, Battelle and NEXUS.
34. K. Cotter, T. J. Drabik, and R. J. Dillon, "Ferroelectric-liquid-crystal/silicon-integrated-circuit spatial light modulator,"*Optics Letters*, **15** (1990), 291.
35. Tarui, T. Hirai, K. Teramoto, H. Koike, K Nagashima, "Application of the ferroelectric materials to ULSI memories," *Applied Surface Science*, **114** (1997), pp. 656–663.
36. M. Gorbatkin, D. B. Poker, R. L. Rhoades, C. Doughty, L. A. Berry, S. M. Rossnagel, "Cu metallization using a permanent-magnet electron-cyclotron-resonance microwave plasma/sputtering hybrid system," *J. Vac. Sci. and Tech. B*, **14** (1996), pp. 1853–1859.
37. Streiter, T. Gessner, H. Wolf, "Geometrical optimization of multilevel interconnects using Cu and low-k dielectrics," *Microelectronic Engineering*, **33** (1997), pp. 429–436.
38. Zhao, M. A. Biberger, V. Hoffman, S. Q. Wang, P. K. Vasudev, T. E. Seidel, "A novel subhalf μm Al-Cu via plug interconnect using low dielectric-constant material as inter-level dielectric," *IEEE Electron Dev. Lett.* **18** (1997), pp. 57–59.
39. T. Tseng, Y. T. Hsieh, C. F. Lin, M. S. Tsai, M. S. Feng, "Chemical-mechanical polishing and material characteristics of plasma-enhanced chemically vapor-deposited fluorinated oxide thin-film," *J. Electrochem. Soc.* **144** (1997), pp. 1100–1106.
40. T. Tseng, Y. T. Hsieh, C. F. Lin, M. S. Tsai, M. S. Feng, "CMP of fluorinated silicon dioxide: Is it necessary and feasible?," *Solid-State Technology,* (1997), pp. 61–66.
41. T. Bohr, "Interconnect scaling - The real limiter to high performance ULSI," *IEDM '95*, (1995), pp. 241–244.
42. Homma, K. Kusukawa, T. Furusawa, H. Miyazaki, N. Kobayashi, "An approach to realize deep submicron multilevel interconnection using low-resistance and low-ε materials," *Electronics and Communications in Japan, Part 2*. **78** (10), (1995), pp. 52–64.
43. Shibata, "Issues and perspective in multilevel interconnection technology for quarter-μm generation," *Electronics and Communications in Japan, Part 2*. **78**(10), (1995), pp. 85–97.
44. M. Sze, *Semiconductor Sensors*, John Wiley and Sons (1994).

45. Kanno, "Uncooled infrared focal plane array having 128x128 thermopile detector elements," *Proc. SPIE* **2269** (1994), pp. 450–459.
46. Markandey, R. J. Gove, "Digital-display systems based on the digital micromirror device," *SMPTE Journal*, **104** (1995), pp. 680–685.
47. M. Monk, R. O. Gale, "The digital micromirror device for projection display," *Microelectronic Engineering*, **27** (1995), pp. 489–493.
48. Semancik, R. E. Cavicchi, K. G. Kreider, J. S. Suehle, P. Chaparala, "Selected-area deposition of multiple active films for conductometric microsensor arrays," *8th International Conference on Solid-State Sensors and Actuators and Eurosensors IX. Digest of Technical Papers*, **1** (1995), pp. 831–834.
49. B. Apte, F. S. A. Sandejas, W. C. Banyai, D. M. Bloom, "Deformable grating light valves for high resolution displays," *Tech. Digest 1994 Sol.-State Sensor and Actuator Workshop*, (1994), pp. 1–6.
50. Spangler, C. J. Kemp, "ISAAC : Integrated silicon automotive accelerometer," *Sensors and Actuators A* **54** (1996), pp. 523–529.
51. Igarashi, "Special features of automotive onboard sensors," *8th International Conference on Solid-State Sensors and Actuators and Eurosensors IX. Digest of Technical Papers*, (1995), pp. 898–899.
52. Vieten, P. Weber, M. Lapin, "Specialized accelerometers for automotive noise and vibration measurements," *Sound and Vibration*, **30** (1996), pp. 6–7.
53. "Control of vehicle dynamics," *Automotive Engineering*, **103** (1995), 87.
54. "Countering an unexpected skid," *Road and Track*, **46** (1995), 141.
55. "Future vehicle stability systems," *Motor Trend*, **47** (1995), 116.
56. Sains, "Mercedes moves beyond ABS," *Design News*, **50** (1995), 28.
57. T. Howe, B. E. Boser, A. P. Pisano, "Polysilicon integrated microsystems : technologies and applications," *Sensors and Actuators A*, **56** (1996), pp. 167–177.
58. Analog Devices, Datasheet: "ADXL150/ADXL250 rev.0" Norwood, MA (1996).
59. Kuehnel and S. Sherman, *Sensors and Actuators A*, **45** (1994), pp. 7–16.
60. C. Beatty, "A chronology of thermal ink-jet structures," *Tech. Digest 1996 Sol.-State Sensor and Actuator Workshop*, (1996), pp. 200–204.
61. K.D. Wise, J.B. Angell and A.Starr, "An Integrated-Circuit Approach to Extracellular Microelectrodes," *IEEE Transactions on Biomedical Engineering*, **BME-17**, (1970), pp. 238–246.
62. K.D. Wise and J.B. Angell, "A Low Capacitance Multielectrode Probe for use in Extracellular Neurophysiology," *IEEE Transactions on Biomedical Engineering*, **BME-22**, (1975), pp. 212–219.
63. S.L. BeMent, K.D. Wise, D.J. Anderson, K. Najafi, and K.L. Drake, "Solid-State Electrodes for Multichannel Multiplexed Intracortical

Neuronal Recording," *IEEE Transactions on Biomedical Engineering*, **BME-33**, (1986), pp. 230–241.

64. K. Najafi and K.D. Wise, "An Implantable Multielectrode Array with On-Chip Signal Processing," *IEEE Journal of Solid-State Circuits*, **SC-21**, (1986), pp. 1035–1044.
65. K.L. Drake, K.D. Wise,J. Farraye, D.J. Anderson, and S.L. BeMent, "Performance of Planar Multisite Microprobes in Recording Extracellular Single-Unit Intracortical Activity," *IEEE Transactions on Biomedical Engineering*, **35**, (1988), pp. 719–732.
66. K. Najafi and K.D. Wise, "A Scaled Electronically-Configurable Multichannel Recording Array," *Sensors and Actuators*, **A21-A23**, (1990), pp. 589–591.
67. M. Peckerar, S.A.Shamma, M. Rebbert, J. Kosakowski, and P. Isaacson, "Passive Microelectrode Arrays for Recording of Neural Signals: A Simplified Fabrication Process," *Review of Science and Instrumentation*, **62**, (1991), pp. 2276–2280.
68. W.G. Regehr, J. Pine, and D.B. Rutledge, "A Long-Term *In Vitro* Silicon-Based Microelectrode-Neuron Connection," *IEEE Transactions on Biomedical Engineering*, **BME-35**, (1988), pp. 1023–1032.
69. N.A. Blum, B.G. Carkhuff, H.K. Charles, R.L. Edwards,and R.A. Meyer, "Multisite Microprobes for Neural Recordings," *IEEE Transactions on Engineering*, **38**, (1991), pp. 68–74.
70. P.R. Kennedy, "The Cone Electrode: A Long-Term Electrode that Records from Neurites Grown onto its Recording Surface," *Journal of Neuroscience Methods*, **29**, (1989), pp. 181–193.
71. A.B. Frazier, D.P. O'Brien, and M.G. Allen, "A Two Dimensional Metallic Micro Electrode Array," Proceeding of the 1993 Micro Electro Mechanical Workshop, (1993), pp. 195–200.
72. G.E. Loeb, W.B. Marks, and P.G. Beatty, "Analysis and Microelectronic Design of Tubular Electrode Arrays Intended for Chronic, Multiple Single-Unit Recording from Captured Nerve Fibres," *Medical & Biological Engineering & Computing*, **15**, (1977), pp. 195–201.
73. T. Akin and K. Najafi, "A Micromachined Silicon Sieve Electrode for Nerve Regeneration Applications," *1991 International Conference on Solid-State Sensors and Actuators*, (1991), pp. 128–131.
74. S. C. Terry, J. H. Jerman, and J. B. Angell, "A gas chromatographic air analyzer fabricated on a silicone wafer," *IEEE Trans. Electron. Devices*, **ED-26**, (1979), 1880.
75. R. R. Reston and E. S. Kolesar, "Silicon-micromachined gas chromatography system used to separate and detect ammonia and nitrogen dioxide--part I: design, fabrication, and integration of the gas chromatography system," *IEEE J. Microelectromech. Sys.*, **3**, (1994), 134.

76. R. R. Reston and E. S. Kolesar, "Silicon-micromachined gas chromatography system used to separate and detect ammonia and nitrogen dioxide—part II: evaluation, analysis, and theoretical moedling of the gas chromatograph system," *J. Microelectromech.* Sys. **3**, (1994), 147.
77. G. Ocvirk, E. Verpoorte, A. Manz, H. M. Widmer, "Integration of a micro liquid chromatography onto a silicon chip," in *Proc. Tranducers 95*, (1995).
78. A. Manz, Y. Miyahara, J. Miura, Y. Watanabe, H. Miyagi, and K. Sato, *Sensors & Actuators*, **B1**, (1990), 249.
79. D. E. Raymond, A. Manz, and H. M. Widmer, "Continuous sample preparation using free-flow electrophoresis on a silicon microstructure," in *Proc. Transducers 95*, (1995).
80. A. Manz, "Miniaturized chemical analysis systems based on electroosmotic flow," in *IEEE Micro Electro Mechanical Systems Conf.*, (1997).
81. C. S. Effenhauser, A. Manz, and M. H. Widmer, "Glass chips for high-speed capillary electrophoresis separation with submicrometer plate heights," *Anal. Chem.*, **65**, (1993), 2637.
82. Z. H. Fan and D. J. Harrison, "Micromachining of capillary electrophoresis injectors and separators on glass chips and evaluation of flow at capillary interconnections," *Anal. Chem.*, (1994), p. 66, 177.
83. S. C. Jacobsen, A. W. Moore, and J. M. Ramsey, "Fused quartz substrates for microchip electrophoresis," *Anal. Chem.*, **67**, (1995), 2059.
84. H. Nakanishi, T. Nishimoto, N. Nakamura, S. Nagamachi, A. Arai, Y. Iwata, and Y. Mito, "Fabrication of electrophoresis devices on a quartz and glass substrates using a bonding with HF solution," in *IEEE Micro Electro Mechanical Systems Conf.*, (1997).
85. B. K. Gale, A. B. Frazier, and K. Caldwell, "Micromachined electrical field-flow fractionation (m-EFFF) system," in *IEEE Micro Electro Mechanical Systems Conf.*, (1997).
86. M. A. Northrup, C. Gonzalez, D. Hadley, R. F. Hills, P. Landre, S. Lehew, R. Saiki, J. J. Sninsky, and R. Watson, "A MEMS-based miniature DNA analysis system," in *Proc. Transducers 95*, (1995).
87. P. Wilding, M. A. Shoffner, and L. J. Kircka, "Manipulation and flow of biological fluids in straight channels micromachined in silicon," *Clin. Chem*, **40**, (1994), 1815.
88. M. Albin, R. Kowallis, E. Picozza, Y. Raysberg, C. Sloan, E. Winn-Deen, T. Woudenberg, and J. Zupfer, "Micromachiningand microgenetics, what are they and where do they work together?" in *IEEE Solid-State Sensor and Actuator Workshop*, (1996).
89. R. C. Anderson, G. J. Bogdan, and R. J. Lipshutz, "Miniaturized genetic-analysis system," in *IEEE Solid-State Sensor and Actuator Workshop*, (1996).

90. R. Austin, R. Carlson, S. Chan, J. Knight, "Micro and Nanopore Structurres for Biological Applications," *Proceeding of the 1997 International Conference on Solid-State Sensors and Actuators,* (1997), pp. 1303–1305.
91. R. Carlson, C. Gabel, S. Chan, R. Austin, J. Brody, and J. Winkleman, "Self-sorting of White Blood Cells in a Lattice," *Physical Review Letters*, (1997).
92. H. Ayliffe, R. Rabbitt, P. Tresco, and B. Frazier, "Micromachined Cellular 33 Characterization System for Studying the Biomechanics of Individual Cells," *Proceeding of the 1997 International Conference on Solid-State Sensors and Actuators*, (1997), pp. 1307–1310.
93. E.T. Den Braber, et al., "Effects of Parallel Surface Microgrooves and Surface Energy on Cell Growth," *Biomedical Materials Research*, **29**, (1995), pp. 511–518.
94. L. Bousse, "Whole Cell Biosensors," *Proceeding of the 1995 International Conference on Solid-State Sensors and Actuators*, (1995), pp.483–486.
95. P.E. Anderson, P.C.H. Li, R. Smith, R.J., Szarka, and D.J. Harrison, "Biological Cell Assays on an Electrokinetic Microchip," *Proceeding of the 1997 International Conference on Solid-State Sensors and Actuators*, (1997), pp. 1311–1314.
96. A.B. Frazier, "Micromachined Precision Cutting Tools," *17th International Conference of the IEEE Engineering in Medicine and Biology Society*, (1995).
97. D. Mathieson, U. Beerschwinger, B.J. Robertson, S.J. Yang, R. L. Reuben, R.A. Lawes, F.N. Goodall, J. Spencer, and R.S. Bartholomew, "Design Considerations for Surgical Micro-fluid Actuators," *Nanobiology*, **3**, (1994), pp. 123–132.
98. D. Mathieson, B.J. Robertson, U. Beerschwinger, S.J. Yang, R. L. Reuben, A.J. Addlesee, "Micro-Torque Measurements for a Prototype Turbine," *Journal of Micromechanics and Microengineering*, **4**, (1994), pp. 190–196.
99. Abbott Critical Care Systems, "Our Products Guide," Salt Lake City, UT.
100. T. Bell, K.D. Wise, and D. Anderson, "A Flexible Micromachined Electrode Array for a Cochlear Prosthesis," *Proceeding of the 1997 International Conference on Solid-State Sensors and Actuators*, (1997), pp. 1315–1318.
101. R.A. Boie, "Capacitive Impedance Readout Tactile Image Sensor," *Proceedings of the International Conference on Robotics*, (1984), pp. 370–378.
102. M.H. Raibert, "An All-Digital VLSI Tactile Array Sensor," *Proceedings of the International Conference on Robotics*, (1984), pp. 314–319.
103. S. Sugiyama, K. Kawahata, M. Yoneda, and I. Igarashi, "Tactile Image Detection using a 1K-Element Silicon Pressure Sensor Array," *Sensors and Actuators*, **A21-A23**, (1990), pp. 397–400.

104. R.J. DeSouza and K.D. Wise, "A High Density Tactile Imager for Reading Embossed Characters," *Proceedings of the 1994 Solid-State Sensor and Actuator Workshop*, (1994), pp. 13–16.
105. K. Frank and M.C. Becker, "Microelectrodes for Recording and Stimulation," *Physical Techniques in Biological Sciences*, **5**, (1964).
106. J. Skrzypek and E. Keller, "Manufacture of Metal Microelectrodes with the Scanning Electron Microscope," *IEEE Transactions on Biomedical Engineering*, **BME-22**, (1975), pp. 435–438.
107. E. Petersen, "Silicon as a mechanical material," *Proc. IEEE,* **70** (1982), pp. 420–457.
108. Robbins and B. Schwartz, "Chemical etching of silicon, I. The System HF, HNO_3 and H_2O," *J. Electrochem. Soc.*, **106**, (1959), 505.
109. Robbins and B. Schwartz, "Chemical etching of silicon, II. The System HF, HNO_3 and $HC_2H_3O_2$," *J. Electrochem. Soc.*, **107**, (1960), 108.
110. Robbins and B. Schwartz, "Chemical etching of silicon, III: A temperature study in the acid system," *J. Electrochem. Soc.*, **108**, (1961), 365.
111. K. Kendall, "On etching very narrow grooves in silicon," *Appl. Phys. Lett.*, **26** (1975), 195.
112. Bean, "Anisotropic etching of silicon," *IEEE Trans. Electron. Dev.*, **ED-25**, (1978), pp. 1185–1193.
113. Bassous, "Fabrication of novel three-dimensional microstructures by the anisotropic etching of (100) silicon," *IEEE Trans. Electron. Dev.*, **ED-25**, (1978), 1178.
114. For example, Motorola's MPX10 and MPX12 silicon piezoresistive pressure sensors.
115. Delpoux, "Motorola and MEMS : The way up to a surface micromachined accelerometer," *Microelectronics Journal*, **28** (1997), pp. 381–387.
116. J. Sniegowski, S. L. Miller, G. F. LaVigne, M. S. Rodgers and P. J. McWhorter, "Monolithic geared-mechanisms driven by a polysilicon surface-micromachined on-chip electrostatic microengine," *Tech. Digest 1996 Sol.-State Sensor and Actuator Workshop*, (1996), pp. 178–182.
117. L-S. Fan, Y-C Tai and R. S. Muller, "Integrated movable micromechanical structures for sensors and actuators," *IEEE Trans. Electron. Dev.*, **35**, (1988), pp. 724–730.
118. W. Becker, W. Ehrfeld, P. Hagmann, A. Maner, D. Munchmeyer, "Fabrication of microstructures with high aspect ratios and great structural heights by synchrotron radiation lithography, galvanoformung and plastic moulding (LIGA Process)," *Microelectronic Engineering*, **4**, (1986), pp. 33–56.
119. G. Keller and M. Ferrari, *Tech. Digest 1994 Solid-State Sensor and Actuator Workshop*, (1994), pp. 132–137.
120. G. Keller and R. T. Howe, "Nickel-filled Hexsil thermally actuated tweezers," *8th International Conference on Solid-State Sensors and*

Actuators and Eurosensors IX. Digest of Technical Papers, (1995), pp. 376–379.

121. Smith, S. Montague, J. Sniegowski, C. Barron, J. Allen, B. Davies, J. Fleming, J. Murray, P. McWhorter, R. Howe, B. Boser, A. Pisano, M. Lemkin, T. Roessig, W. Clark, and T. Juneau, "Micromachined sensor systems on a chip: The integration of MEMS with CMOS and its applications," *Proc. 43rd International Instrumentation Symposium,* (1997), pp. 199–206.
122. Smith, J. Allen, S. Montague, and J. Ellis, "Critical issues for the application of integrated MEMS/CMOS technologies to inertial measurement units," *Proc. SPIE* **3046**, (1997), pp. 242–247.
123. Smith, S. Montague, J. Sniegowski, J. Murray R. P. Manginell, and P. McWhorter, "Characterization of the embedded micromechanical device approach to the monolithic integration of MEMS with CMOS," *Proc. SPIE* **2879**, (1996), pp. 306–314.
124. Yun, R. Howe, and P. Gray, "Surface micromachined, digitally force-balanced accelerometer with integrated CMOS detection circuitry," *Proc. of the IEEE Solid-State Sensor and Actuator Workshop '92*, (1992), 126.
125. Howe, "Polysilicon integrated microsystems: technologies and applications," *Proc. Transducers '95*, (1995), pp. 43–46.
126. H. Smith, S. Montague, J. J. Sniegowski, J. R. Murray, and P. J. McWhorter, "Embedded micromechanical devices for the monolithic integration of MEMS with CMOS," *IEDM Tech. Digest*, (1995), pp. 609–612.
127. H. Smith, S. Montague, and J. J. Sniegowski, "Material and processing issues for the monolithic integration of microelectronics with surface-micromachined polysilicon sensors and actuators," *Proc. SPIE* **2639**, (1995), pp. 64–73.

CHAPTER 2

Wet Chemical Etching of Silicon and SiO_2, and Ten Challenges for Micromachiners

Don L. Kendall
Robert A. Shoultz*
University of New Mexico, Center for High Technology Materials

**Also with Superkinetic Inc.*

CONTENTS

2.1 Perspective / 43

2.2 Isotropic Etching / 44
  2.2.1 The HNA and porous silicon systems / 45
  2.2.2 Wafer cleaning / 49
  2.2.3 Hydrogen passivation and surface roughness / 51
  2.2.4 The strange {100}symmetry / 51
  2.2.5 Defect delineation etches and the Sirtl CrO_3:HF family / 54

2.3 Anisotropic Etching / 57
  2.3.1 KOH / 57
  2.3.2 Other metal hydroxides: NaOH, LiOH, CsOH, RbOH / 65
  2.3.3 EDA, EDP, TMAH, NH_4OH, N_2H_4, and Gallates / 66

2.4 Additives for KOH: alcohols, triethanolamine, and metals / 68

2.5 Anisotropic Etching Mechanisms / 69
  2.5.1 Oxidation-dissolution model / 69
  2.5.2 Hydration model / 69
  2.5.3 Electrochemical model / 70

2.6 Etching of SiO_2 / 71

2.7 Five Wet Etching Challenges / 75
  2.7.1 Unidentified zones in the HNA and Sirtl ternaries / 75
    2.7.1.1 Etching for atomic step and Miller index delineation / 76
  2.7.2 How do pits relate to interface states on Si? / 78
  2.7.3 Major questions from hydroxide etching mechanisms / 78
    2.7.3.1 Why do organic bases etch {100} faster than {110}? / 78
    2.7.3.2 Can the "hydroxonium" mechanism be proven? / 79

2.7.3.3 Why is there large trace metal effect of KOH on Si? / 79
2.7.3.4 Determining the maximum anisotropy obtainable / 80
2.7.4 Narrow vertical grooves on the (100)? / 81
2.7.5 Why the large EDTA effect of KOH etching of SiO_2? / 81

2.8 Five New Micromachining Challenges / 82
2.8.1 Ordered bubble flows as analogs for superconductivity / 82
2.8.2 One-way flow in porous Si using tethered buckyballs / 83
2.8.3 Deuterium isotope separation on {111}Si / 84
2.8.4 Vertical semipermeable membrane structures / 85
2.8.5 Directional-facing atomic steps for chiral boxes / 87

2.9 Conclusion / 88

Acknowledgments / 89

References / 89

2.1 PERSPECTIVE

Wet chemistry is still very important in semiconductor and micromachining technology, in spite of important progress being made by the dry (or plasma) etching processes. For example, wet chemistry is used in the majority of the process steps employed in the production of integrated circuits. This is primarily because the cleaning of Si-processed wafers by wet chemical means is more effective than plasma techniques. Nevertheless, the dry methods are often favored where submicron geometries and narrow grooves must be etched into dielectrics and Si for integrated circuits (ICs), for example, in the fabrication of trench capacitors. The lateral etching generally observed using wet chemistry is a severe limiting factor in producing submicron-width features.

This chapter is not a comprehensive review. Rather, we intend it as a guide for today's work and also to pose significant challenges for future research and development. We emphasize those areas where wet etching has a significant advantage over dry etching. This often falls into the rapidly growing area of micromachined sensors and other applications, but it is also still quite important in integrated circuit technology.

There will be two groups of readers of this work. The first group, "the users," will be interested in the five challenges of the primary chemical etching systems for Si and oxides. They will find new ways of thinking about the generally isotropic etchants for Si, namely, the hydrofluoric, nitric, and acetic acid or HNA system, as well as the HF-CrO_3-water defect delineation family. They will also find an extensive discussion of the anisotropic etch, KOH:water, including the mechanisms and idiosyncracies of this system and other hydroxide systems on both Si and SiO_2.

The second group, "the perusers," will be more interested in where this field of research could lead in the field of micromachining. We also suggest five rather speculative, even revolutionary, structural challenges derived from our experience in micromachining that have not yet been explored in any recent publications of which we are aware. This second group of readers may wish to turn to Sections 2.7 and 2.8 to get an idea of where we are going, and then come back to the earlier sections to understand where these concepts come from.

We believe that wet chemical etching will probably be used ultimately for making trench capacitors when and if the {110} wafer orientation is adopted as a significant player in the manufacture of complementary metal-oxide-semiconductor (CMOS) circuits. The latter will almost certainly occur when low-temperature ICs become important in the world of high-temperature superconductors and low-temperature CMOS logic. This is because the twofold mobility advantage of holes on the {110} along certain azimuthal directions (relative to {100} wafers) at 77 K allows much more balanced CMOS circuits than the {100} surface orientation.[1] This change in wafer orientation will also allow much higher aspect ratio grooves (greater than 100:1) to be economically etched for trench capacitors. We do not believe that dry etching will be able to do this uniformly across many large wafers for many years, although there have been significant advances in this application,[2] with aspect ratios of 50:1 now possible. These high ratios using reactive ion etching (RIE) require multiple steps of etching and chemical vapor deposition (CVD) of polymers on the groove walls.[2] However, it is the mobility advantage of holes on the {110} that will ultimately be the deciding factor in such a decision for low-temperature CMOS. Incidentally, the somewhat higher interface charge state density of the {110} than that on the {100} is not a serious problem on modern ICs because of the very thin gate oxides, which give a high oxide capacitance and hence a very small contribution to the threshold voltage.

We first briefly discuss some of the main examples of isotropic etching and wafer cleaning, and also the defect delineation systems. These data are readily available in other sources and are not covered in great detail here. The exception to this will occur when we can show the reader how to develop his or her own etching method for a particular purpose or we can add something that is not well covered in the literature.

We also give a collection of useful data for anisotropic etching, which is most generally useful for bulk micromachining rather than surface micromachining. ("Surface for show, bulk for dough." It will be interesting to see how long the fast-charging surface micromachinists will take to reverse this situation.) In particular, we provide an extensive discussion and useful etch rate information in simple engineering notation for the KOH etching of Si (and SiO_2) at all temperatures and concentrations. This etchant has become the workhorse for many micromachining applications because of its simplicity and relatively benign effects on the environment. We also discuss the benefits of important additives to the KOH system, as well as other hydroxide and organic base types of anisotropic etchants.

2.2 Isotropic Etching

The workhorse system for the isotropic wet etching of Si is certainly the so-called HNA system. This system has been well characterized[3] and is usually represented by a three-phase diagram showing contour plots of the etch rates at various places in the three-component field. We discuss important regions of this three-phase diagram and some methods for designing an etch for a particular function. We also treat wafer cleaning, hydrogen passivation, microroughness, the unusual symmetry of the {100} surface, and the various defect delineation etches, since Si etching is often involved in each of these subjects.

2.2.1 The HNA and Porous Silicon Systems

The dissolution process in HNA is almost universally accepted to be oxidation of the Si by the nitric acid, followed by dissolution of the oxide by the HF. The acetic acid acts as a buffer and in many applications can be replaced by water where the buffering is not essential. Some of the more interesting special cases are listed in Table 2.1, several of which come from reviews by Runyan and Bean[4] and Kern and Deckert[5]. These special cases are also shown in Fig. 2.1 on the ternary diagram. For a brief overview of etching mechanisms during isotropic etching and the cleaning of Si, we especially recommend Kern and Deckert[5].

TABLE 2.1 Special HNA etches for Si, where the volume percentages are listed in the order of the standard bottle concentrations of HF(49%), HNO_3(70%), acetic acid (100%).

Etchant	Composition/(vol%)	Etch Rate/Side at 25°C Comments
CP4A[3,4] (Chem. polish #4A)	3:5:3 (27:46:27)	16 µm/min
PLE (Planar etch)	2:15:5 (9:68:23)	2 to 6 µm/min Unstirred to ultrasonic stir.
CDE[6] (Concentration dependent etch)	1:3:8 (8:25:67)	2µm/min for $R < 0.001$ Ω cm; 0.001 µm/min, $R > 0.1$ Ω cm.
ACDE[7] (Alternate concentration dependent etch)	1:40:20 (2:65:33)	For concentration profiling.
SSE (Si slow etch[8])	1:99:0 (1:99:0)	2 µm/min in small windows at first, then 0.5 µm/min at long times (see text and Fig. 2.2)
SHF (Si HF Etch[9])	100:0:0 (100:0:0)	0.3 A/min. Dilute HF and water are faster than concentrated!
STE (Stain etch[9])	1000:1:0 (99.9:0.1)	Very slow, say < 10 A/min.
JSE (Junction stain etch[10])	1:3:20 (4:13:83)	

The etch rates (at 25°C) in Table 2.1 and in Fig. 2.1 are given for one-sided etching, even though Refs. 3–5 often give the rate for two sides since the original application for such etches was for overall slice thinning. In modern technology, one side is often protected by a patterned nitride, thick oxide, or metal protective layer.

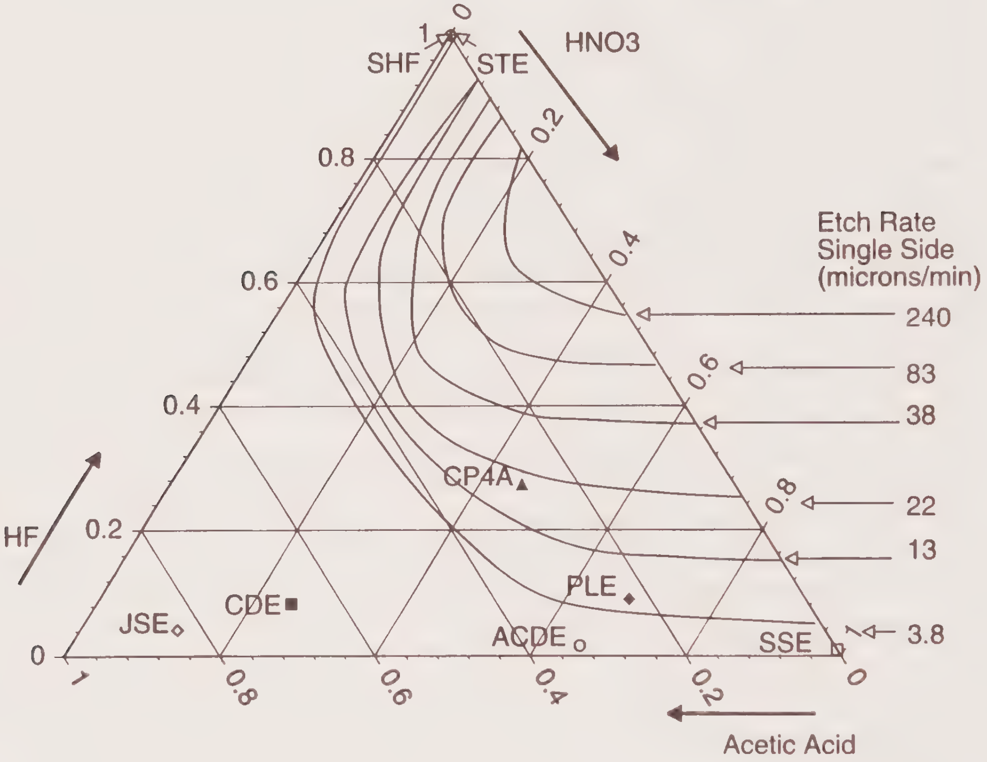


Fig. 2.1. The HF, nitric, acetic (HNA) ternary diagram showing the position of a few well-known etchants, as well as contour lines for the etch rates for a single side.

The etch rates given in Table 2.1 are only rough approximations since the rates are affected by the number and positions (even vertical versus horizontal disposition) of wafers in the container, as well as the density and shapes of the patterned etch areas. The lower rates are generally typical of single wafers in a large unstirred container, and the high rates can be obtained with small openings in large protected regions, with vigorous stirring, or with closely spaced wafers. The latter situation causes stirring of the etchant by the neighbors and also contributes etch products that can accelerate the dissolution process. The etch products can contribute to a process called "auto-catalytic etching," which can considerably speed up the chemical reactions. This is especially important in the interesting concentration-dependent etch[6] (CDE) proposed by Muraoka et al., who used certain additives to suppress the excess etching due to the etch products. This etchant dissolves heavily doped Si of both $n+$ and $p+$ type about

2000 times faster than lightly doped Si, but only under conditions where the low- and high-concentration wafers are well separated, or the etch products are chemically compensated, or swept away in a large container.

The large effect of stirring almost all of the above etchants is due primarily to the fact that the reaction diffusion boundary region is much smaller under intense stirring conditions. This means that the reactants can diffuse across the smaller boundary during stirring more readily by Fick's first law, and hence the reaction can proceed more rapidly. The same general effect occurs in many other diffusion-limited situations, as will be discussed for other etchants later, such as KOH at low concentrations. However, the HNA etchants, at least at room temperature and above, are almost always isotropic at all useful concentrations; that is, they etch at almost the same rate in all crystallographic directions. By contrast, the alkaline etchants such as KOH become almost completely reaction-rate limited at high concentrations and it turns out that they are extremely dependent on crystal direction.

The very rapid CP4A (chemical polish #4A) and the much slower planar etchants (PLE) both result in wafer edges that are thinner than the center. This is again a diffusion limitation effect since more reactants are available at the edge of the wafer. Planar is a good polishing etchant if one can tolerate the rounded edges. However, 50 to 100 μm must be removed from each side of a lapped wafer in a tilted rotating beaker before the surfaces are smooth enough for even medium-geometry lithography (say, 5-μm minimum geometry). Submicron geometry requires chem-mechanically polished wafers.

A similar edge-tapering effect occurs for the Si slow etch (SSE) (see Table 2.1), which is mostly nitric acid with only 1% added HF, and no acetic acid. An additional effect has been observed in this etchant, namely, the Si etching in narrow openings in oxides (both vertically and laterally) is much faster in the initial stages, as shown in Fig. 2.2. In this figure, the oxide etch rate in this etchant is only about 8 nm/min at 25°C, so a thermal oxide is an adequate etch mask for many purposes. For still slower Si etch rates, these rates can be reduced proportionally by diluting the HNO_3 still further. For example, 1:1000:0 should give rates about 10 times slower.

An additional effect during etching in SSE is observed across a nitride etch window in a vertically disposed wafer.[11] The upper part of the opening is subject to much greater turbulence (hence greater mixing and deeper etching near the top edge) than the bottom of the opening. This extra etching at the edges of an opening can be made more uniform by etching the wafers in a horizontal position, but the effect will still be present because of the greater availability of reactants near a nonetching mask edge.

Hoelke[12] has recently shown that the HNA system can become anisotropic if the etch rate is reduced by lowering the temperature to about -30°C. He notes that circular windows in a nitride mask result in six-sided depressions in {111} and {110} wafers when etched at this low temperature in mixtures similar to the CP4A mixture.[12] This anisotropy manifests itself at the lower temperature because the viscosity is greatly increased (and hence the diffusion coefficients of the reactants are much reduced). Thus, the surface reaction rates on the different crystal orientations dominate the dissolution process, with the {111} rates being the slowest, which is similar to the strongly anisotropic etchants to be discussed later.

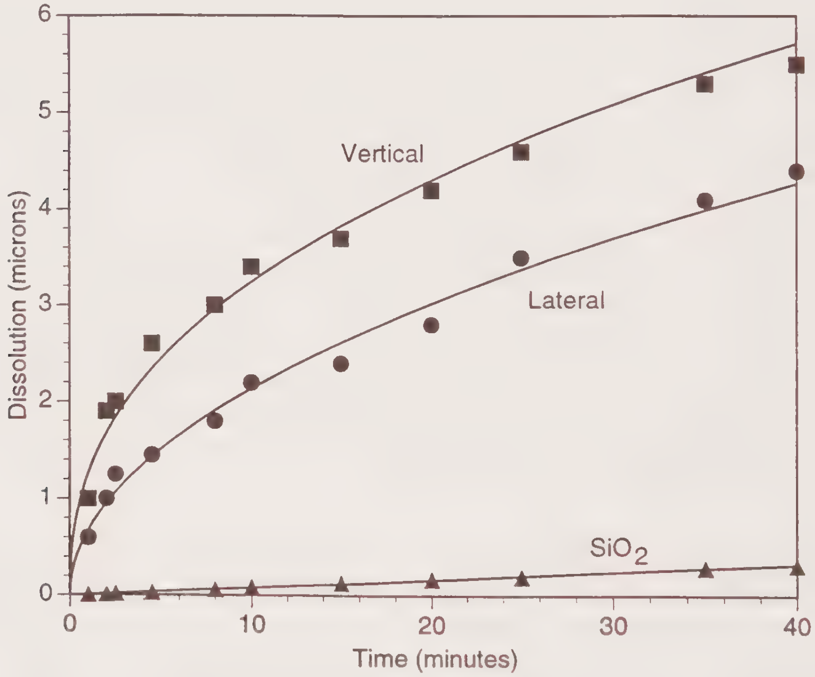


Fig. 2.2. Etch rates in a 12-µm window at 25°C for 1 vol% HF in nitric acid (the SSE etch of Table 2.1 and Figure 2.1). The rate on SiO_2 is also given.

Finally, we should comment on the opposite extreme (from mostly nitric) of mostly HF with a few added drops of HNO_3, which we have specified as 1000:1:0 as a rough estimate. This will begin to etch when the nitric and exceeds a certain value, but the first visible effect is a severe stain on the surface that is not soluble in HF. This stain (STE in Table 2.1) is sensitive to carrier concentration and conductivity type and is often used as a quick check on epitaxial or diffusion-layer thickness. However, this method is not highly reliable since too much nitric acid will cause the Si to etch at a finite and uncontrollable rate. Similar comments could be made about the junction stain etch (JSE) of

Table 2.1, but JSE is probably more selective than STE in its staining properties for different conductivity types.

This is a good time to remind the reader that Si can also be dissolved very slowly in pure HF at 0.3 A/min, as well as even faster in dilute HF:water and buffered HF mixtures[9] at about 0.8 A/min. This is called the Si-HF, or SHF, etch in Table 2.1. It is interesting that extrapolation of these results to pH 7 for pure water predicts an etch rate of about 1.5 A/min at 25°C. We mention this later in our discussion of very dilute KOH etching.

Still another form of "etching" of Si can be done in pure or dilute HF by making the Si the anode and using a counterelectrode of, say, Pt to form "porous silicon." This introduces very small pores into the Si (as small as 1 nm and as large as about 3 μm, depending on the resistivity and type of the Si and the current density).[13] There is a gap between about 0.1 and 1 μm in the pore diameters that can be produced.[14] This porous Si layer can then be oxidized rapidly to produce Si on insulator (SOI), structures.[15] The layer can also be etched very rapidly in dilute KOH or NaOH and used as a sacrificial layer in certain surface micromachining applications.[16]

The pores above generally follow a torturous path and result in very low-density layers, with porosities over a wide range from 0.2 to 0.9, depending on the etching conditions. However, incredibly straight and deep pores (several hundred micrometers deep with diameters around 2 μm) can be produced in {100} by introducing the pores from the front side of lightly doped *n*-Si and flooding the back of the wafer with nonpenetrating light. This unusual scheme allows the minority carriers (holes) to diffuse to the front of the wafer, where they meet the tips of the relatively large pores that have already started. Thus, the tips collect all the anodization current and the pores advance in straight lines through the whole wafer.[17] To make these pores uniform in spacing, the desired geometry is controlled by preetching regular pits using an anisotropic etchant such as KOH:water. We could write much more about porous Si and its applications, but Ref. 13 provides an excellent review.

2.2.2 Wafer Cleaning

The almost universal standard for Si wafer cleaning is the so-called RCA cleanup,[18] which consists of an SC1 and an SC2 cleaning process, both at about 70°C for about 10 min. SC1 is a very efficient remover of particles and consists of a 1:1:5 mixture by volume of ammonium hydroxide (NH_4OH), peroxide (H_2O_2), and water. It is also sometimes called APM, for ammonium-peroxide mixture.[9]

It has recently been found that a 0.25:1:5 mixture of APM is equally effective and results in smoother surfaces on an atomic scale[20]. A still more dilute mixture of 0.1:1:5 is also adequate for removing particles, but requires a longer soak time of over 30 min. The more dilute solutions are generally desirable for environmental reasons. A Megasonic[21] vibrator (about 900 kHz) is quite helpful during the SC1 step, whether diluted or not, since the vibration helps loosen the particles from the Si surface, where the particles are held by strong electrochemical and/or van der Waals forces.[21]

A major problem of the SC1 mixtures is that metal contamination from the containers, the starting wafers, or from the chemicals themselves often ends up on the Si surfaces. Typical Fe surface concentrations after an SC1 treatment, for example, are $2.5\times10^{12}/\mathrm{cm}^2$, where $1\times10^{11}/\mathrm{cm}^2$ is considered the maximum allowable Fe surface density in a clean process.[19] A desirable solution would be to somehow modify the SC1 process so that the metals would not be deposited on the Si surface. Various labs are working on such solutions, but until they are successful, the original SC2 mixture is generally used to remove the relatively high metal concentrations that are left after the SC1 treatment.

The SC2 mixture consists of a volume ratio of 1:1:5 of $HCl:H_2O_2$:water and is also typically applied for about 10 min at 70°C. It is sometimes called HPM, or hydrochloric-peroxide mixture.[19] It is very successful in removing the Fe and several other metals from the Si surface, often lowering the Fe surface density[19,20] to about $7\times10^{10}/\mathrm{cm}^2$.

The SC1 and SC1 processes are often preceded by a so-called piranha (or Caro`s acid) treatment at 95°C (or higher temperature) for 10 min in a sulfuric acid:peroxide mixture, or an SPM solution.[19] This mixture varies somewhat around a 3:1 mixture of these chemicals and the exothermic reaction of mixing is often used to reach the desired temperature. It is not very good for removing either particles or metals, but is good for removing organic contaminants that might otherwise compromise the subsequent SC1 and SC2 treatments.

The final treatment after the SC2 process is a dip in dilute HF to remove any oxide on the surface. This dilute HF dip leaves the surface hydrophobic and highly vulnerable to picking up any particles from the air or from the final DI rinse.[22] On the good side, however, the HF dip followed by a DI rinse also leaves a hydrogen-passivated surface that greatly impedes the formation of the native oxide, often extending the time to form one monolayer of oxide to more than 16 hr.[23] Omitting the DI rinse after the HF dip is even more effective in this regard,[24] although the remnant HF may cause other problems in some applications. It has also recently been discovered by Ohmi et al. that an Si surface after any HF treatment adsorbs several monolayers of water,[24] which is quite similar to the retention of water molecules on freshly cleaned stainless steel surfaces.[25] This is a problem for scanning tunneling microscopy (STM) analysis in some cases, but is not a problem in many other applications.

The buffered oxide etch (BOE) is also sometimes used for the final cleanup, but BOE is more often used to remove oxide in oxide openings during lithographic processing. The most common version of BOE consists of a so-called 7:1 ratio that is designed to give a convenient etch rate for thermal oxide of 0.1 μm/min at 25°C.[4] It consists of 7 volume parts of 40 wt% $NH_4F:H_2O$ and 1 volume part of 49% HF. The buffering apparently helps stabilize the pH under an oxide ledge and therefore causes less lateral undercutting of the oxide mask than does an HF:water mix with the same oxide etch rate.[26]

2.2.3 Hydrogen Passivation and Surface Roughness

Another treatment that can result in an extremely well H-passivated surface on {111} wafers[27] is "pH modified BOE" in which BOE means buffered oxide etch as discussed in the previous section. This unusual mixture is obtained by adding NH_4OH to BOE until a pH of about 9 is reached. This results in a slow anisotropic etchant that etches misalignment steps rapidly, but the {111} surface etches much more slowly. The net result on a well aligned {111} wafer is an almost perfect atomic {111} flat and very efficient H-passivation. This should also be an excellent final etch after producing vertical {111} studs on the {110} surface using KOH, as will be discussed in a later section.

Incidentally, this pH-modified BOE is not necessarily a good etchant for producing flat surfaces on the {100} since this orientation tends to form multiple {111} facets in isotropic etchants, thereby causing a roughened surface on an atomic basis.[27] Similar remarks could probably be made about the {110}, but we have not seen any reports on this surface using this etchant. As discussed later, there are etchants (ethylenediamine, EDA, and EDA with pyrocatechol, EDP) that etch the {110} quite slowly relative to the {100} so such atomic flatness might also be produced on the {110} surface under certain conditions.

2.2.4 The Strange {100} Symmetry

The {100} is an interesting surface in itself in that even chem-mechanical polishing of perfectly oriented (100) wafers does not result in perfectly atomically flat surfaces, except in local domains about 100 nm in extent.[28] This is the result of the diamond crystal structure of Si, which is formed from two interpenetrating face-centered cubic (FCC) lattices separated by one quarter of the lattice constant along the *x*, *y*, and *z* axes. If one imagines that one FCC lattice is red and the other is blue, and one manages to cleave the crystal exactly across the red lattice on the (100) plane, then one would notice that the surface atoms would have a sort of twofold symmetry, with the two dangling electron bonds on each surface atom all facing, say east and west. Then, if one carefully removes one monolayer of the red atoms from the surface, the next layer of atoms would be in the blue FCC lattice and shifted one-quarter of a lattice distance in x, y, and z. The surface atoms would now all be blue and the two dangling bonds on the surface atoms would all be facing north and south, so a twofold symmetry would again be observed on the surface. However, now the bonds face in the direction perpendicular to the first cleaved case.

The actual situation is even more complicated than described because the neighboring bonds tend to attract each other and neighboring atoms on either FCC lattice move toward each other, and restructure the surface to form rows of paired surface atoms. Each pair in the row is called a dimer. For the red atom surface above, the dimers would form in rows that run north and south, even though the bonds on this surface originally faced east and west before restructuring. This unique situation exists in one form or another on alternate layers of Si. So on any real polished surface, the local deviations from atomic flatness will cause "domains" of N-S dimer rows interspersed with domains of E-W dimer rows. [The average structure of such real surfaces is fourfold since there will be equal numbers of the two types of domains on a well aligned (100) surface.] These effects are generally observed under ultrahigh vacuum (UHV) conditions using various surface measurement techniques.

If the above surface is even slightly misaligned from the perfect (100), then monolayer atomic steps can be observed, with each step having lines of dimers either parallel or perpendicular to the step risers. This is shown in Fig. 2.3, which was taken with an STM of a well-annealed sample misaligned by only 0.25° toward the [110] direction.[28,29] For reasons that are related to the different directions of the dimers on alternate steps as discussed in the previous paragraph, alternate step edges are straight and jagged.[28]

When the surface is misaligned from the (100) by more than about 1.5°, the steps often tend to merge into each other and form double steps, and the terraces on the steps all have the same dimer orientation. For example, during molecular beam epitaxy (MBE) of slightly misaligned {100} Si, the dimer rows all tend to be parallel with the step edges.[30] By contrast, during sputtering there is a tendency for such steps to have most of the dimers perpendicular to the step edges.[29]

Thus, we see that there is a complex set of effects associated with this domain structure near the {100} surface. We mention the effect here because there are almost no data on the effect of chemical etchants on the properties of these domains, either in well-aligned or misaligned {100} surfaces. One can imagine that subsequent oxide or other dielectric deposition could well be affected by the starting surface morphology. For example, one can only speculate on the effects on the threshold voltage of depositing a thin oxide or other dielectric on the surface of Fig. 2.3 compared with the partially annealed chaotic ("intermediate state") sputtered surface shown in Fig. 2.4. Similar comments should apply to the Schottky barrier heights of metals carefully deposited on two such extreme morphologies. The variations among surface morphologies at the atomic level of different chemical treatments may span an even larger range than the two shown here. The following section will give a greater appreciation of the different surface characteristics obtainable with a range of etchants.

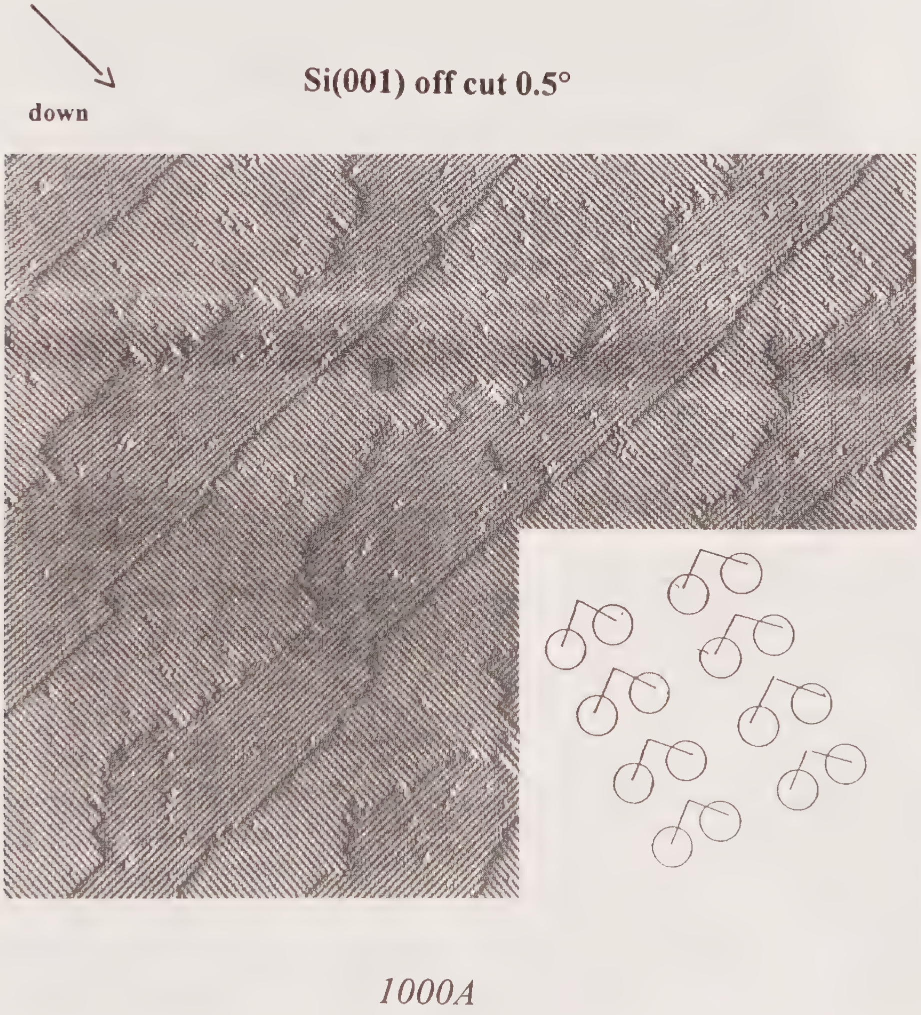


Fig. 2.3. Single atomic steps on the (100) of Si as portrayed by scanning tunneling microscopy. The sample was annealed at 1200°C in an ultrahigh vacuum and then cooled to 25°C for measurement in the UHV.[29]

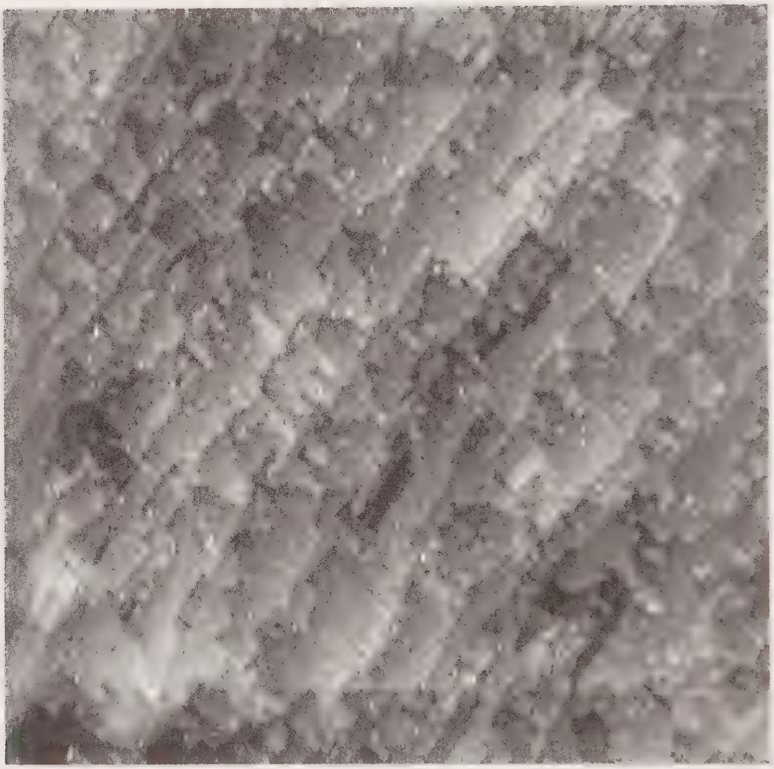


Fig. 2.4. Highly cluttered steps after sputtering the sample of Fig. 2.3 for 60 min at 465°C and then annealing for 10 min at 600°C.

2.2.5 Defect Delineation Etches and the Sirtl CrO_3:HF Family

Runyan and Bean[4] have provided a good summary of the defect-delineating etches, except for the Seiter etch discussed later. One of the earliest was the 1:3:10 HNA mixture of Dash,[31] which is close to the concentration-dependent etch at 1:3:8 (see Fig. 2.1), but the Dash etch was used primarily to show etch pits on the {111}. The CDE aspect certainly existed in the Dash etch, but was not usually emphasized. After the early work of Dash, the dislocation delineators became dominated by HF plus CrO_3 in varying proportions for different crystal orientations; Sirtl[32] for {111}, Schimmel[33] for {100}, and Yang[34] for all orientations, including the {110}.

The final member of this family is that of Seiter[35] for a mixture that leaves mounds instead of pits so it can be used as an "integrational etchant." This "mounding" does not disappear with additional etching and allows one to count the dislocations in a volume of Si rather than only delineating defects near the surface and then losing identification of each layer of pit defects as the etching proceeds. The Seiter etch contains much more CrO_3 than the other etchants, which apparently accounts for the drastically different behavior. The volume compositions of each are as follows, where we have chosen a normalized HF concentration to facilitate comparisons.

Sirtl	1 HF,1(5 M-CrO_3)	for the (111);
Yang	1 HF,1(1.5 M-CrO_3)	for the (110), (and other hkls)
Schimmel	1 HF,0.5(1 M-CrO_3)	for the (100);
Seiter	1 HF,9(12 M-CrO_3)	for mounds instead of pits.

We note again that the nomenclature is such that the Sirtl etch consists of 1 part by volume of HF from the bottle (of 49% solution) and 1 part by volume of a 5-molal solution of CrO_3 in water.

We have also prepared Fig. 2.5 to show a graphical view of the differences among these various defect delineating etches. The first three etchants are in the northwest-central zone and the Seiter is near the south central region, where we have placed the oxidizing chemical in the southeast corner, and HF at the top (north). The Sirtl etch is the strongest oxidizing solution (containing the most CrO_3) of the three pit etches, which apparently is optimum for producing pits on the (111) surface.[34] The Yang mixture is a good compromise solution that produces reasonable pits on all of the major planes (without stirring). Yang points out that the optimum pit-producing solutions are ordered by the oxidizing strengths and also by their surface energy on the different planes, with the {111} > {110} > {100}. We note here that this is also the same order as the decreasing surface state density[36] on these three orientations. We return to this in Sections 2.7.1 and 2.7.2.

Also in Fig. 2.5, we see that the Seiter etch with its high oxidizing power and weaker HF concentration is a quite distinct etchant from the pit formers. One could speculate that the Cr tends to deposit on the Si much more readily since the etchant is closer to the solubility limit of CrO_3 in water at room temperature (solubility of 16.5 molal, or at 82.5% of our arbitrarily established 20 molal at 100%). Thus, atomic disparities or defects on the surface may tend to nucleate Cr or some chromic oxide in specific regions of surface energy. Such a deposition appears to block the etching of the surface and leave mounds.

On the above basis, we also propose that different crystal orientations may etch quite differently in all regions of the ternary of Fig. 2.5, even in the absence of defects. For example, a misoriented (100) surface has been shown to have misalignment steps as in Fig. 2.3, but most chemical etches do not delineate these steps very well. Similarly, the (114) and (5 5 12) surfaces of Si have been shown to yield surfaces with atomic ridges with near-perfect regular spacings of about 2.0 nm.[37,38] We have recently been studying the etching and other properties of such unusual surfaces.[39]

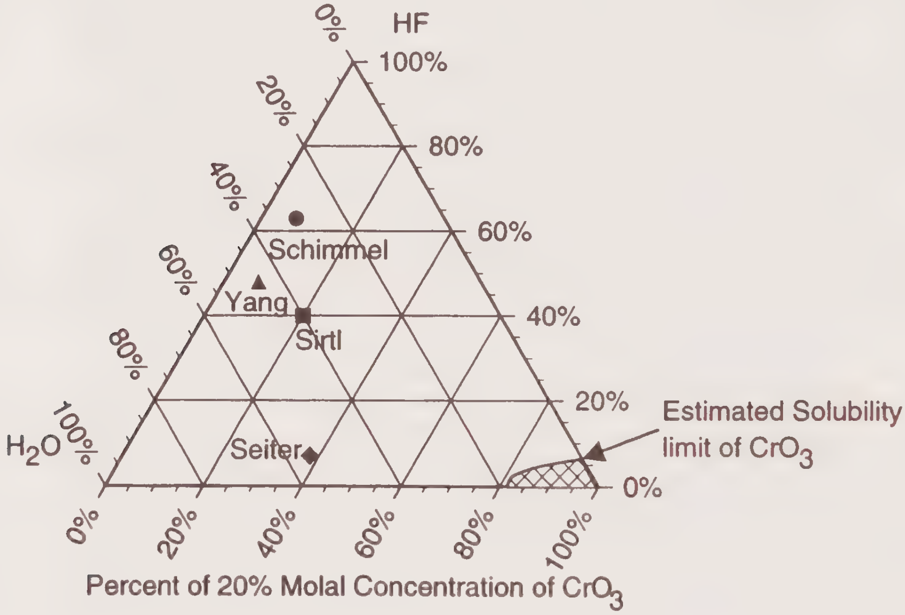


Fig. 2.5. The artificial volume ternary of the Sirtl family defect delineators. The "volume" of the dry powder CrO_3 is arbitrarily given relative to a 20 molal concentration. The hatched area delineates the estimated region where the CrO_3 concentration exceeds the solubility limit, which is 16.5 molal in pure water (about 82.5% of 20 molal on the bottom ordinate). See the text for a sample calculation.

An etchant related to the above group is the Secco etch,[40] which consists of 2HF,1(0.15 M-$K_2Cr_2O_7$). This is quite good at delineating oxygen-induced stacking faults on the {100}.

A final etchant that is often used is the Wright–Jenkins (or the "Wright") etch,[41] which is a 2:1:2 HNA mixture (see Fig. 2.1) with added water and Cu and Cr additives, namely, 60 ml HF, 30 ml HNO_3, 60 ml acetic acid, 30 ml(5 M-CrO_3), 2 g $Cu(NO_3)_2$, and 60 ml of water. This is a true etchant for Si, and the metal additives give it good selectivity for delineating defects on both the {111} and {100}. The original references should be consulted for detailed instructions for using all of these interesting etchants.

2.3 ANISOTROPIC ETCHING

A large number of alkaline-type etchants have been found to be anisotropic, that is, there is a strong crystallographic dependence on etch rates. An occasional reference is made to an acidic system that has anisotropic character, such as CP4A at very low temperatures mentioned earlier,[12] or NH_4F doped with Fe, Co, or Cu ions,[42] but we will emphasize the alkaline systems since they are the ones most often used in modern micromachining, as well as in V-MOS integrated circuits and in certain types of Si-on-insulator structures. The first systems studied were based on KOH and organic mixtures of various types, as well as hydrazine, and later ethylenediamine or EDA with pyrocatechol added as an accelerator. The EDA and hydrazine both make strong bases by clipping H^+ ions off of two water molecules to release large concentrations of OH^- ions. They each have advantages over KOH and NaOH and several other metallic hydroxides in their slow etching of SiO_2, which is often used as an etch mask. However, they do not generally have crystallographic etch ratios as high as the metallic hydroxides. They are also much more toxic and must be used with extreme care. Their other properties will be discussed later after a summary of the known effects of the various simpler hydroxides.

In this paper we summarize the literature on all the alkaline-based etchants. We also present some new data on KOH etching that relates to aging and contamination of the solutions, as well as more information on the use of additives for special applications, and a newly reported method to get smooth etched surfaces using multiple-frequency ultrasonic vibration.

2.3.1 KOH

The KOH:water system has become the major etchant used for micromachining applications. It was discussed in some detail in a recent review[43] and some items will be repeated here from a somewhat different viewpoint. Previous reviews should also be consulted for specific information that we refer to briefly.[44–47]

The interpolated etch rates of the {100} and {110} are shown in Figs. 2.6 and 2.7 in intervals of 10°C. This is a new presentation of the data in a previous paper[46] with some account taken of other work.[47,48] The calculated curves come from a reanalysis of the data[46–48] at different temperatures and KOH concentrations in terms of the hydration model of Glembocki et al.[49] An empirical fit that gives results very similar to the Glembocki model is given by the following equations, which are easily programmed on small calculators. These equations are accurate to about 20% over the full range of temperature and KOH concentrations.

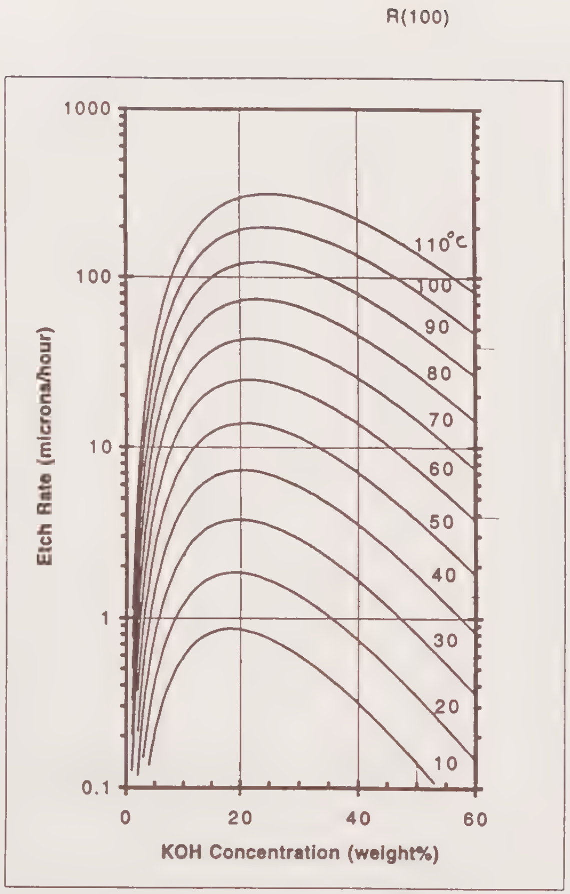


Fig. 2.6. Empirically calculated etch rates for {100} Si in KOH:water using Eq. (1a) of text.

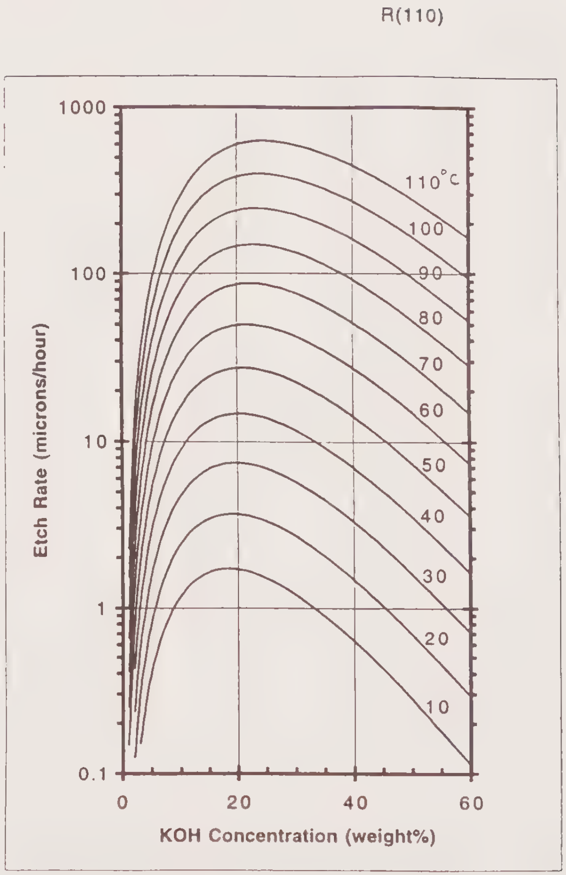


Fig. 2.7. Empirically calculated etch rates for {110} Si in KOH:water using Eq. (1b) of text.

$$R_{100} = 2.6\times10^{6}\ W^{2.5}\ \exp\text{-}(w/300 + 0.48)/k(T{+}273)\ (\mu\text{m/hr}) \tag{1a}$$

$$R_{110} = 5.2\times10^{6}\ W^{2.5}\ \exp\ \text{-}(w/300 + 0.48)/k(\mathrm{T}{+}273)\ (\mu\text{m/hr}) \tag{1b}$$

where W is the actual wt% of KOH in water (as monitored by specific-gravity measurements, not the nominal wt% of KOH pellets, which is typically off by 15 to 20% due to the water in the pellets), k is Boltzmann's constant of 8.617×10^{-5} eV/^{o}K, and T is the temperature in degrees Celsius. We note that the rates on the {110} are forced to be exactly 2.0 times those on the {100}. (The actual enhancement factor[46] was 2.03 ± 0.20 over a range of concentrations from 5 to 55 wt% and temperatures from 30 to 85°C.)

In the above analysis, we have also taken account of a set of data taken at 40°C in Germany without any stirring using a European source of chemicals. The purpose is to show possible differences in etch rates due to the presence of different impurities in the KOH, and also to show the effect of aging of the solution in Pyrex containers for 1 and 3 months. We return to this important point in Section 2.7.3.3. The original etch rates are as much as 30% slower than those taken using an American source of KOH (with a mild stirring action by continual rocking of the container). However, as seen in Fig. 2.8, after aging in Pyrex and being opened to the air a few times, the German etch rates are very similar to the American rates. It is not clear if this is evidence that there is more or less carbon dioxide in the European KOH, since both claimed to be in the 0.5 to 1.0% range of this impurity in the KOH pellets.

Another possibility is that the heavy metal concentration is somewhat less in the fresh European KOH, but that it became more or less equivalent to the American source after leaching some metals out of the Pyrex storage container. We note in passing that we have not detected any difference in etch behavior whether done in Pyrex or in polyethylene containers. The enhancement of the etch rate by the presence of a large quantity of the heavy metal Fe is evident in the single data point of Fig. 2.8, where we purposely added 100 ppm to the KOH. In a later experiment[50] with 50 ppm of Fe (added by the chloride in both of our experiments) in 30% KOH:water, we noted a number of cubic-shaped precipitates on the bottom of wide grooves on {100} Si. However, the {111} walls and the general etching of V-grooves appeared to be normal in other respects, except for the enhanced etch rate on the {100}.

A recent article[51] by Bressers et al. also noted a modest increase in etch rate in a 19 wt% KOH mixture when $K_3Fe(CN)_6$ was added in small quantities. However, when larger concentrations were added, the formation of undesirable pyramidal hillocks was suppressed, but the etching stopped in narrow oxide openings below a certain size. When still larger concentrations of the Fe compound were added, the etching stopped completely, even in very large openings. The authors discussed these unusual effects in terms of a sort of "chemical anodization" and the geometry dependence as an "oxygen diffusion effect."

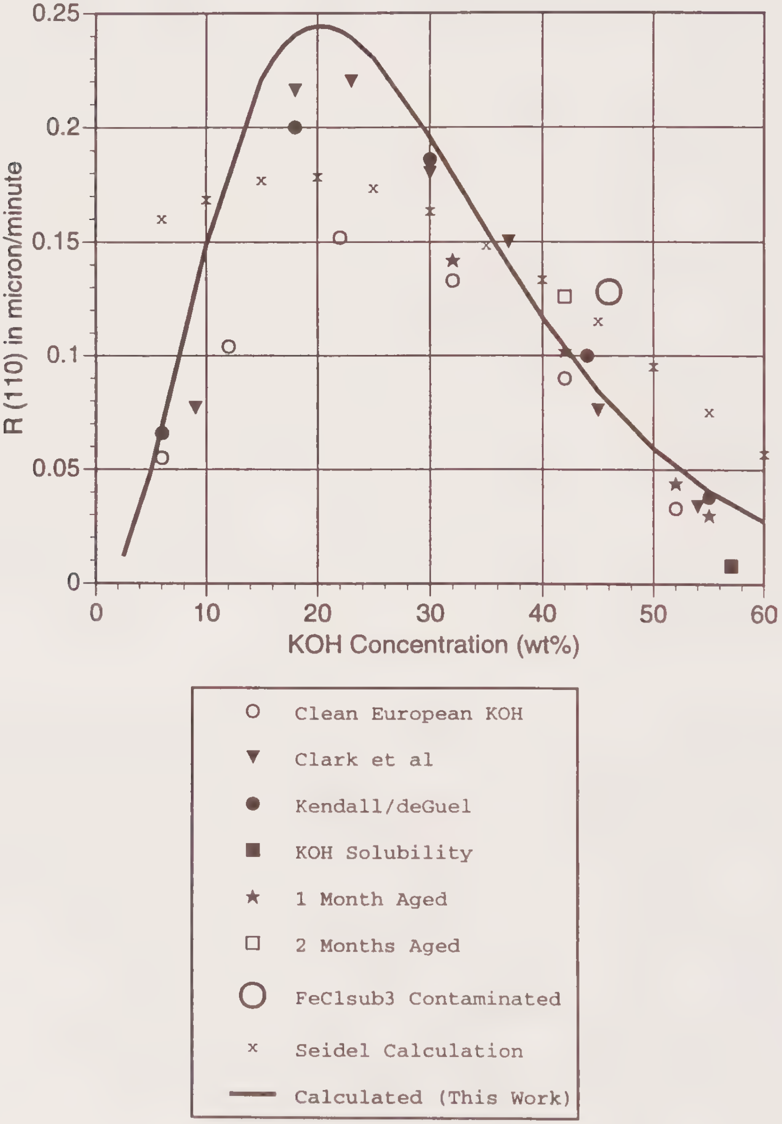


Fig. 2.8. KOH:water etch rates in European and American labs and the effect of aging and contamination with Fe (via the chloride). Also shown is our empirical fit (solid line) and that of Seidel (small xs).

Finally, we have also plotted our empirical estimate of the expected etch rate on the {110} in Fig. 2.8 (solid line), as well as the rate based on Seidel's fourth power of the water model (small x's). Our German values in clean chemicals are closer to the unstirred data at both low and high concentrations, but Seidel's estimate is closer in the 20 to 30% range. Higher etch rates seem to be typical of most KOH solutions under normal laboratory conditions. This is clear from the American KOH data of Kendall and de Guel (solid squares),[46] and also the 40°C data of Clark and Edell (solid circles).[48] We note that the empirical prediction is quite consistent with all the data at high KOH concentrations (except for the fresh and clean European KOH), but is about 10 to 15% too high in the range of 20 wt% of KOH. These experimental values are within the measurement uncertainty.

The important point from Fig. 2.8 is that one should not expect great reproducibility in KOH etch rates, unless the chemicals are fresh and the heavy metals like Cu and Fe are at least consistent in concentration (see also Section 2.7.3.3). We do not have any good information on the effect of carbon dioxide in the solution. However, we know that a KOH solution exposed to normal air will continually pick up CO_2. This can be avoided by etching in an N_2 atmosphere, and the CO_2 can also be reduced in the solution by bubbling N_2 through the solution in a closed system. Our experience with such N_2 experiments has not shown any great changes in the etching behavior, but we did not measure the rates very carefully.

If we compare the rates predicted by Eqs. (1a) and (1b) with Seidel's work,[47] we find that our values on the {100} are lower by 5 to 20% near the peak rates at all temperatures, and they are typically higher than Seidel's rates on the {110} by about the same amount. By contrast, our experimental values are generally within about ±10% of Clark and Edell's values for the {110} at all temperatures and concentrations except at 10 wt%, where their values are as much as 50% higher. This latter discrepancy is probably due to stronger stirring in their experiments, as explained in the following discussion.

The much higher predicted values at 10% and below by Seidel may also be due to greater stirring, but we believe these high values and the markedly higher values at high KOH concentrations are also a consequence of the quite different exponents they chose in what amounts to a hydration model similar to that of Glembocki et al.[49] The latter were able to fit the etch rates rather well with the following chemical reaction.

$$Si + 2OH^- + 2H_2O \rightarrow Si(OH)_2(O^-)_2 + 2H_2 \qquad (2.2)$$

The Glembocki model is based on completely independent experimental data on the number of water molecules that are attached to an OH^- ion over a wide range of KOH concentrations, as shown in Fig. 2.9. (Our normalized calculations using Eq. (1b) mimic the general form of the physics-based Glembocki model.) Seidel et al. use an empirical model that effectively chooses the exponents on the left side of this reaction as 1/2 and 4 (rather than 2 and 2), which we believe incorrectly elevates the predicted values at both low and high KOH concentrations (see Fig. 2.8). Our predictions in Eqs. (1a) and (1b) are also based on experimental data and a particular linear dependence of the activation energy and a power law dependence of the preexponential on KOH concentration. (The activation energy Q for the etch rates at different KOH concentrations W is chosen to vary linearly from 0.48 to 0.68 eV between W of zero and 60 wt%, respectively. The preexponential for the (100) rate, R_0, is fit to a power law $2.6\times10^6\ W^{2.5}$ (µm/hr). We basically force our R_0 values to fit our linear Q assumption. Clark's and Seidel's Q and R_0 values show considerably more scatter, but are of the same magnitude.) As mentioned in previous reports,[43,46] stirring causes a marked increase in etch rates at low KOH concentrations, but has much less effect at high concentrations. This was given as evidence that the etching process at high concentrations of KOH is primarily reaction-rate limited, and each crystallographic surface has its own etch rate. However, the process becomes more and more diffusion-rate limited at low KOH, so stirring has a much larger effect, as discussed earlier with regard to the thickness of the boundary layer in the HNA system.

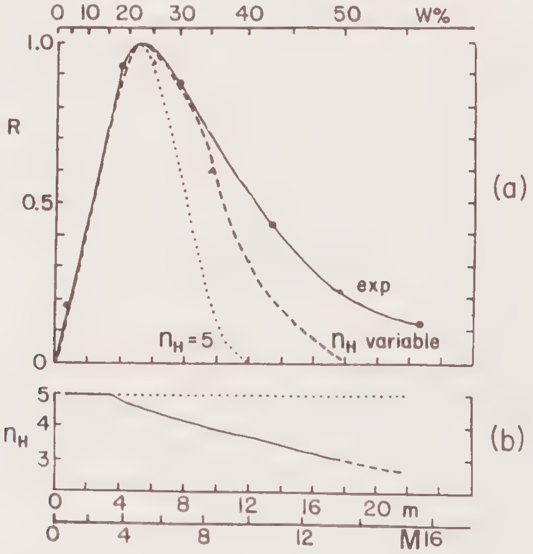


Fig. 2.9. Normalized hydration model at 30°C of Glembocki et al.[49] at different hydration numbers. Our empirical fit mimics the general behavior.

A side effect of this shift from reaction rate limitation at high KOH to a diffusion-limited process at low concentrations is that the etch ratio between the {110} and {111} steadily decreases from perhaps as high as 500:1 at 55 wt% KOH to near 1:1 at 0% KOH(pure water).[46] Thus, to obtain high crystallographic etch ratios, one should use rather high KOH concentrations, even though the Si etch rates will be much lower and the relative oxide (mask) etch rates, compared with Si, will be higher. The {100}:{111} ratios will be only about half of the values quoted above, for the reasons mentioned with regard to Eqs. (1a) and (1b).

We also should note that the highest {110}:{111} etch ratios of about 200 obtained by Clark and Edell[48] are only about half of those obtained in earlier work.[46,52] The reason for this discrepancy probably lies in the different methods used to measure the undercutting on the very slow etching {111} side walls. The differences here are ultimately due to the fact that the very slow etch rate on the perfect {111} is extremely difficult to measure. As discussed in an earlier paper,[53] this measured etch rate depends on correction factors to account for the simultaneous dissolution of the oxide mask, and is probably also affected by stress between the Si and the masking material, whether it is oxide or nitride or whatever.

The etch rate on the slow {111} is also strongly affected by the oxygen content of the crystals and the heat treatment that occurs during growth, as well as the "resistivity stabilization heat treatment" around 700°C that the wafer producers give the crystals to reduce the oxygen donor formation in the 400 to 500°C range. In extreme cases these low-temperature heat treatments can turn a lightly doped *p*-type wafer into an *n*-type wafer.

Of more relevance to anisotropic etching, however, are the oxide precipitates of, say, 200 to 1000 Å in diameter that form during normal wafer processing in the range of 1000 to 1200°C. These large precipitates form on the small oxygen nuclei of 15 to 50 Å diameter that are created during the 700°C treatments.[4] An excellent picture of the effect of a 2-hr 700°C treatment is shown by Freeland et al.,[54] in which an oxygen-containing crystal shows about 10^4 precipitates/cm^2 after an 8-hr treatment at 1200°C, whereas the same 1200°C treatment without the 700°C heat treatment shows absolutely no precipitates.

One of the challenges for Si wafer manufacturers is to produce wafers that are suitable for both high-quality anisotropic etching and that also are adequately stable in terms of resistivity shifts during normal device processing. Recent work at Virginia Semiconductor Inc. (VSI) and MOXTEK Inc. has gone a long way toward addressing these issues.[55] Larson at MOXTEK processed {110} wafers that were specially grown at VSI to contain lower and more uniform oxygen concentrations, and that had no 700°C heat treatment. This procedure eliminated the "oxygen swirl" pattern sometimes seen in the widths of the narrow vertical grooves etched in KOH. Furthermore, a higher oxidation temperature of at least 1100°C to produce the oxide etch mask also reduced the undercutting on the {111} by about 25% compared with 1000°C oxidation. This latter improvement was almost certainly due to the high-temperature dissolution of any remaining precipitates (see Ref. 4, Fig. 8.46). These very deep vertical grooves of about 1000 μm traversed the wafers completely in order to produce X-ray collimators (Soller slits), and as such they provided an excellent test vehicle for the effect of the oxygen precipitation process.

The best data on the effect of heat treatments during crystal growth, as well as subsequent heat treatments after growth, are probably that of Iwasaki et al.[56] This recent paper is full of relevant experiments on the issue of oxygen precipitates as they relate to the so-called gate oxide integrity (GOI) defects. We assert here that these same precipitation GOI "defects" are also responsible for the excess etching that occurs on the {111} walls of grooves etched in KOH:water mixtures. Their experiments are completely consistent with the observations of extra etching at the oxide defects of the previous paragraph,[55] although the focus was quite different in the two studies.

A somewhat earlier paper discussed the smoothness of the {111} surfaces on the side of a V-groove on {100} Si. Float-zoned FZ wafers were generally smoother than CZ wafers after such etching, and longer processing cycles also had a roughening effect on the groove walls.[58] We have not seen a marked difference in the etching properties of the few FZ wafers that we have measured, but it is reasonable that some differences in the {111} etch rates should be present.

Similarly, we would expect that so-called swirl-free CZ crystals should also have lower etch rates on the {111} since the oxygen content is presumably more uniform in such crystals, as well as more resistant to oxide precipitates. Ideally, for minimum etching on the {111}, we would want the oxygen to be well dispersed rather than clustered into precipitates. Such an effect explains the significantly different lateral etching discussed in the VSI-MOXTEK experiments as well as the large variation often observed in different published papers. For example, an early work[52] suggested the etch ratios of {110}:{111} as high as 400:1 could be obtained on swirl-free CZ wafers if they were not given the 700°C heat treatment before the 1100°C oxidation process. However, as discussed earlier,[43,53] the jury is still out on the precise anisotropic ratios using a completely stress-free and etch-resistant mask (see also Section 2.7.3.4 for a discussion of this issue).

An additional treatment could also be applied to both vertical grooves on the {110} and V-grooves on the {100} in order to obtain smoother {111} side walls.

This is the pH-modified BOE process mentioned earlier for producing wide expanses of atomically flat {111} surfaces.[27] While we are speculating on ways to improve the {111} side walls, one might also consider a brief etch in an HNA mixture at very low temperatures (-30°C), as discussed earlier,[12] since the {111} etch rate will be much slower than the other orientations.

We should also mention that the etch rate of the {111} does not go to zero at zero KOH concentration, since the etch rate of Si in pure water is actually finite. This was discussed earlier with regard to etching in dilute HF solutions, and is also treated in the work by Glembocki et al.[49]

When silicon dioxide grown at different temperatures is used as an etch mask against KOH etching, one might expect to see different lateral etch rates in the Si even when the etched grooves are perfectly aligned with the traces of the {111} planes. This would be due to the different densities[59] of high-temperature oxides (above 1150°C) of 2.20 g/cm^3 as compared with the much higher density at low temperatures (below 700°C) of 2.28 g/ cm^3. There have been no careful studies of the oxidation temperature on etch-anisotropy, and the oxide precipitate effects mentioned earlier would have to be eliminated or accounted for to make such experiments meaningful.

A final comment should be made on a possible way to obtain smoother etch bottoms of both {100} and {110} Si when etching with KOH:water mixtures. This involves ultrasonic agitation during KOH etching using sequential millisecond pulses of three different frequencies rather than a single frequency.[60] The smoothness of the large groove bottoms was quite striking using this method. These authors also reported much more uniform etching of very narrow (1 µm) grooves on the {110} using this method. We have not reproduced this work, but this technique would be a valuable adjunct to our normal processes.

2.3.2 Other Metal Hydroxides, NaOH, LiOH, CsOH, and RbOH

There have been studies of all the other metal hydroxides for etching Si, though none has any overwhelming advantage over KOH for most purposes. However, each one has special properties and may find application in particular circumstances. The relative etch rates of Si in the metal hydroxides are probably in the same basic sequence as the rates in SiO_2, namely, NaOH > KOH > LiOH > RbOH > CsOH, at equivalent molar concentrations of, say, 4 M. The relative etch rates for the oxide at this concentration are 8:6:5:3:1, respectively.[61] The Si etch rates have not been studied systematically for this purpose.

NaOH etching of Si has been reported by several authors.[25,31] It is similar in behavior to KOH, but is less selective for oxides and probably less anisotropic under most circumstances. It is also less desirable from a contamination standpoint for metal oxide silicon field effect transistor (MOSFET) production because of the more rapid diffusion of Na relative to other metals in gate oxides. Even ignoring the Na contamination problem, it shares with the other metallic hydroxides the difficulty of purification of the chemicals themselves to better than a few parts per million for impurities such as Fe, Cu, and several other heavy metals. For this and other reasons, the nonmetallic hydroxides discussed in the next section are often preferred etchants for some applications.

CsOH has been found[63] to be generally similar in behavior to KOH. It reaches a peak rate at about 20 wt%, similar to KOH, and is about 30 to 40% slower (rather than 600% slower for SiO_2 mentioned earlier). This latter comparison of relative rates between KOH and CsOH implies that the oxide and pure Si do not depend on the same limiting factors, but that other factors are involved with Si etching, such as atomic steric and hydration effects as discussed earlier. These differences are treated in detail in Section 2.5 on etch mechanisms.

LiOH is also a slower etchant than KOH at the obtainable concentrations.[49] The hydration model of Glembocki et al.[49] predicts a monotonic increase in etch rate (no peak rate with change in concentration) because of the low hydration number for this etchant. The sparse data are not adequate to make a definite statement, but a maximum rate is indeed not apparent up to a 5 M concentration (which is near the solubility limit of LiOH in water).

RbOH has also been studied,[64] but the etch rates were quite variable, perhaps due to impurities in the RbOH similar to those mentioned with regard to KOH (see Fig. 2.8).

2.3.3 EDA, EDP, TMAH, NH_4OH, N_2H_4, and GALLATES

The much greater chemical purity of certain nonmetallic basic solutions, as well as greater etching selectivity relative to oxides and heavily *p*-type Si, has often caused workers to adopt other hydroxides or organic bases as etchants for Si micromachining. The most commonly used of these has been ethylenediamine–pyrocatechol. We discuss them roughly in the order of their importance in the manufacturing world.

Ethylenediamine and its faster etching cohort with pyrocatechol added are still quite commonly used, even though there has been a shift away from them for environmental reasons. The primary advantage is their much improved selectivity of Si etch rate compared with that of SiO_2, being on the order of 100 times more selective than KOH. Thus, the oxide becomes an excellent etch mask without the added complexity of using a nitride or other etch mask.

A secondary advantage is that EDA and EDP are excellent *p*-type etch stops, with the etch rate at a boron concentration of $1 \times 10^{19}/\mathrm{cm}^3$ being about 140 times larger[65] than at $1 \times 10^{20}/\mathrm{cm}^3$. The comparable ratios for KOH[65] at weight percents of 43, 24, and 11% are 10, 17, and 80 at 60°C, so except at low concentrations of KOH (and concomitant rough surfaces and slow etch rates), EDP is a superior p^+ etch-stopping system. Another etchant that is less damaging to one's health that might be thought to be useful for this purpose is trimethyl ammonium hydroxide (TMAH), but the *p*-type etch stop ratio at these two boron concentrations[50,66] is only about 5.

A disadvantage of EDA and EDP is that they have poor anisotropy compared with KOH etchants, being on the order of 25:1 between the {100} and {111}, instead of at least 100:1 for the higher concentrations of KOH and water. The pyrocatechol additive to EDA does not change the anisotropy to any significant extent. It simply acts to accelerate the Si etch rate about a factor of 2, which is an important manufacturing issue.

Tetramethyl ammonium hydroxide is growing in use because of its adequate anisotropy, lower toxicity than EDP, and high selectivity for Si relative to oxide. It is commonly available because of its use as a positive photoresist developer. It also has quite high purity because of its development for the metal–oxide–semiconductor (MOS) business and the fact that it is not one of the metal hydroxides, with the attendant relatively high heavy metal concentrations mentioned earlier. It is not a particularly good *p*-type etch stop, as mentioned earlier with respect to EDP and KOH. Various additives such as H_2O_2 do not improve the *p*-type selectivity of TMAH, but seem to make it worse.[50]

Pyramidal hillocks are a significant irritant during the etching of Si in TMAH,[67] in EDP[68,69] in certain composition ranges, as well as in KOH water concentrations below about 18 wt%. Landsberger et al.[67] make the interesting observation that the hillocks seem to be favored when the {100} etch rate becomes faster than the rate on the {110}. As discussed in some detail by Kendall,[43,53] KOH:water mixtures generally have an etch rate ratio of about 2.0 between {110} and {100}, and almost all the organic hydroxides have ratios less than 1.0. In fact, KOH with even a small amount of isopropyl alcohol tends toward a ratio of less than 1. This latter effect is quite useful in reducing the convex corner etching of KOH:water mixtures.[43,53] More relevant to the hillock formation, KOH:water mixtures above the 18 wt% level are generally free of hillocks, but most of the organic hydroxides appear to have a significant problem in this regard. The greatly suppressed {110} may indeed be the culprit in these solutions, as suggested by Landsberger. Landsberger also discusses the various aspects of how the hillocks form in TMAH, as well as how they may relate to convex corner-etching effects.

Another significant potential benefit of etching in TMAH is the low etch rate of Al when excess Si is added to the TMAH[70,71]. Nevertheless, the Al still etches slightly and care must be taken to avoid overetching the metal.

NH_4OH has been well studied as a cleaning agent for Si, as mentioned earlier, and has recently been considered as an anisotropic etchant.[72] It can be obtained at very high levels of purity. Peroxide additions to NH_4OH at concentrations as low as 0.0033 *M* have been shown to completely stop the etching of Si.[73] This is reminiscent of the "chemical anodization" effect of $K_3Fe(CN)_6$ mentioned earlier with regard to KOH, where the additive apparently blocks etching by causing the formation of a thin oxide film on the surface. Equally surprising, even lower concentrations of peroxide caused the etch rate of Si to be about triple that of pure NH_4OH. This is probably an effect of the oxygen enrichment of the solution, which ultimately leads to the formation of an oxide film at the concentration mentioned earlier. It is quite interesting that the etch rate at 70°C of {100} Si is actually higher in 5 wt% NH_4OH with 0.0025 *M* peroxide added than it is in 5 wt% of the strong base KOH in water, namely, about 0.42 µm/min compared with 0.2 µm/min, respectively. However, the peroxide in NH_4OH has a limited lifetime of about 30 min before dissociation at this temperature.[73]

Hydrazine (N_2H_4) is one of the early anisotropic etchants of Si, and was studied by Lee, who reported[74] reasonable anisotropic etch ratios and very low etch rates for both SiO_2 and Al. A more recent study has been done by Mehregany and Senturia.[75]

Finally, we mention the gallates as etchants to replace the EDP etchants for Si. Their properties are comparable in many respects to EDP, but at much lower levels of toxicity.[76,77]

2.4 ADDITIVES FOR KOH: ALCOHOLS, TRIETHANOLAMINE, AND METALS

The best-known additives to KOH are the alcohols, triethanolamine, and certain trace metals. Various alcohols were added to different hydroxides in the early work of Finne and Klein.[78] These workers also introduced the ethylenediamine–pyrocatechol system, which was discussed earlier. They measured the anisotropy on the major planes of Si using several specific alcohols. However, none of these has been adopted with any regularity except EDP and isopropyl alcohol (IPA) added to KOH. The latter was discussed earlier with regard to the greatly reduced lateral undercutting at convex corners. This alcohol and others have sometimes been said to give smoother etching, but we have not found them to be much better than simple KOH:water at high concentrations.

Triethanolamine (TEA) is another interesting additive for KOH. For example, about 5 vol% of TEA added to various mixtures of KOH:water and 1-propanol can eliminate the {311} facets at the edges of square grooves on the {110}. Such a mixture has been used to produce square cross-section echelon diffraction gratings.[79] This additive also helps flatten the bottoms of wide grooves on {100} Si.

The last additives we mention here are pure metals such as Fe, Zn, and other metals, as well as the chlorides of these metals. We have already discussed the rather large acceleration of the etch rate on the {110} in KOH solutions when the chloride of Fe was added (see Fig. 2.8). We will discuss in some detail the addition of a number of pure metals in Section 2.7.3.3. The most striking of these is probably Zn, which at 500 ppm appears to reduce the etch rate of the {111} by a factor of 2. This is a very surprising discovery and would appear to allow much deeper and narrower grooves on the {110}. However, we cannot be sure that the etch rate on the {110} has not also been reduced, so we must await further measurements.

2.5 ANISOTROPIC ETCHING MECHANISMS

2.5.1 Oxidation-Dissolution Model

One of the earliest ideas to explain the extreme differences between the very low etching {111} planes and the much faster {110} and {100} planes of Si in KOH:water solutions was a "working hypothesis" by Kendall.[52] He suggested that the {111} planes oxidized much faster than the other crystal surfaces in these liquids. This rapid oxidation of the {111} blocked the etching of this closely packed plane, but the other crystal faces did not oxidize fast enough to passivate the surfaces against dissolution. This early idea was consistent with the etching data at 85°C, which showed that the etch rates of thermal SiO_2 and the {111} of Si were quite similar. This suggested that etching of the oxide that formed on the {111} in the KOH:water solution was the rate-limiting step for the dissolution of this crystal plane. However, such a model would predict a much slower etch rate on the {111} than actually observed at room temperature. For this reason a quite different model was described next.

2.5.2 Hydration Model

Kendall[54] and Glembocki et al.[49] suggested that water hydration of the OH^- ion combined with a sort of geometrical, or steric, hindrance of the {111} could explain the very large difference in etch rates among the various crystal planes. The model also provided an excellent quantitative explanation for the simple observation that the etch rate of Si on the {100} and {110} has a distinct maximum at about 22 wt% KOH and the fact that the etch rate falls to near zero near the solubility limit of KOH in water. We emphasize here that the hydration model of Glembocki et al. is based on Reaction (2) and literature values of chemical activity of OH^- ions and water molecules (the reactants). It has no adjustable parameters, yet it fits the data quite well on the {100} and {110} surfaces, as shown in Fig. 2.9. We again reiterate that our empirical model of Eqs. (1a) and (1b) for the etch rate on these planes relies only on a variable preexponential (power law on the weight percent of KOH), and a linear change in activation energy. It is true that it appears to have a shape similar to the data (Fig. 2.9), but it is more of an empirical-thermodynamic model than the first-principle model of Glembocki et al. Also, as mentioned earlier, the empirical model of Seidel based on the fourth power of water fits the data rather poorly at both high and low KOH concentrations (see Fig. 2.8).

The extremely slow etch rate on the {111} apparently has no maximum etch rate at 22 wt% so the {111} cannot be treated by the chemical activity data in the same way as the other major planes. The much slower rate (at all KOH concentrations) on the {111} was attributed to the three water molecules that surround the OH^- ion. This is similar to the so-called hydronium ion of acid solutions, which consists of one or more water molecules attached to the H^+ ions. However, in the hydroxide case, these tripod-shaped $3H_2O{:}OH^-$ (or $H_7O_4^-$) "hydroxonium" complexes straddle almost every dangling bond on the {111} surface and block the etching from Reaction (2) since the free OH^- ions and the few remaining water molecules are walled off from the Si surface.

From Fig. 2.9 we see that the predicted hydration number drops from 5 to about 3 on going from low to high KOH concentrations, and as this drop in hydration number occurs, the {111} etch rate gets slower and slower, which leads to higher and higher ratios between the {110} and the {111}. This etch ratio may become as high as 400 or even higher at high KOH concentrations.[46] We should mention that the maximum ratio has been reported to be as low as 200:1 by Clark and Edell.[48] Various possibilities to explain this difference were discussed in some detail in Refs. 43 and 53, but we now believe that the major effect may be the differences discussed earlier relative to oxygen precipitation defects as a result of different heat treatmentsof the ingot and the wafers.[55] A secondary effect may be the stress between the oxide and the Si.

2.5.3 Electrochemical Model

The last model to be discussed is the electrochemical model of Seidel et al.[47] These authors proposed that the first event in the etching process "is an oxidation process where four hydroxide ions react with one surface Si atom, leading to the injection of four electrons into the conduction band." In the reduction that follows, "the injected electrons react with water molecules to form new hydroxide ions and hydrogen." To explain the large anisotropy between planes, they proposed that "the anisotropic behavior is due to small differences of the energy levels of the back bond surface states as a function of the crystal orientation." The electrochemical model has merit if one is trying to understand the surface electrochemistry, but it needs a more precise geometrical explanation of the "energy levels of the back bond surface states" before it can be used to explain the extreme anisotropy of etching.[43]

Either the hydration model or the empirical models of Seidel et al.[47] and this work [Eqs. (1a) and (1b)] can be adjusted to give reasonable fits to the experimental data, but as shown in Fig. 2.8, the Seidel model involving the water to the inverse fourth power (and OH^- to the 0.25 power) does not fit the data very well at either low or high KOH concentrations. We would not expect it to do so since Glembocki's first principle model involving the inverse second power of free water and the second power of OH^- fit the data quite well (see Fig. 2.9).

Finally, we should note that the maximum etch rates on the fast-etching planes of Si, as well as the maximum in the etch rate of SiO_2 discussed in the next section, are determined by the external qualities of the etchant, such as chemical activity of the reactants, rather than the atomic details (steric) of the surfaces. Nevertheless, it is the atomic details that determine the actual rates of dissolution of the various Si surfaces. We honor this general concept when we apply a simple multiplying factor of 2× at all KOH concentrations to the {110} rate in Eq. (1b) that is 2× the {100} rate of Eq. (1a). The only case where we deviate from this is in the case of the {111}, where the threefold symmetric steric qualities of the surface and the hydration properties of the OH^- ion and the water molecules appear to dominate the etching at all concentrations. This leads to the observation that the relative rates of the fast planes compared with the very slow {111} plane increase more or less continuously as the KOH concentration increases.[46]

2.6 ETCHING OF SiO_2

The wet chemical etching of the various forms of SiO_2 is very well reported in the literature,[80] and we restrict this section to a discussion of sacrificial oxide etching in HF solutions by Monk et al.[81] and a complete analysis and new empirical representation of SiO_2 etching in KOH solutions from four different groups. These two subjects are quite important to the burgeoning fields of surface and bulk micromachining.

Before discussing the etching of sacrificial oxides,[81] we comment on the review by Runyan and Bean[80] of the etching of thermally grown SiO_2 on Si substrates.

They do not make any distinction between oxides grown at different temperatures and conditions (e.g., wet vs. dry). This is a rather good assumption for the etch rates of various thermal oxides, but because of the different densities[59] and film stresses of these oxides, it may not be valid for the lateral etching of {111} Si underneath the oxide, as discussed earlier. Runyan and Bean show that buffered oxide etch dissolves thermal oxide almost twice as fast as simple dilute HF at equivalent molarities.[80] As mentioned earlier,[26] the BOE also leads to markedly less undercutting (lateral etching) of the oxide under a photoresist layer.

The most common BOE composition is the so-called 7:1 mixture, which consists of 7 volume parts of 40 wt% NH_4F in water to 1 volume part of 49% HF. This gives an etch rate of 100 nm/min at room temperature, which is convenient for normal processing. It should be compared with 1.8 µm/min for the etch rate in undiluted bottled 49% HF.

There are typically two approaches to enhancing the lateral etch rates for clearing the oxides underneath polysilicon or silicon nitride layers hundreds of micrometers wide in surface micromachined devices. The first and most significant method would add P (or B and P) to the oxide to make phosphosilicate glass (PSG) or boron-phosphosilicate glass (BPSG). This can increase the etch rate by a factor of 100 to 1000 for dilute HF mixtures, depending on the amount of P or B added to the oxide. The second method allows an additional factor of 2 to 3 in the etch rate to be obtained by adding up to 50 vol% of HCl to the dilute HF. The latter mixture has a significantly lower pH than a simple HF mixture and can actually decrease the etch rate on polysilicon and silicon nitride, thereby improving the selectivity of the oxide etching for the important structural materials often involved in micromachined parts.

Another common etchant for single-crystal Si and amorphous SiO_2 is KOH at various concentrations. Extensive data have been published for the etch rates of thermal oxides at various temperatures and KOH concentrations.[46–48,50,61] We have reexamined all of this work and developed an empirical relation that gives values accurate to about 20% over a very wide range of both temperature and KOH concentration. The equation is the following:

$$R = 2.2\times10^{9}\, W\,(1.5\times10^{-4}\, W^{2.15} + 1)\,\exp\text{-}(0.795 + 6\times10^{-6}\, W^{2.5})/k(T+273) \qquad (3)$$

where R is in micrometers per hour, W is the wt% of KOH in water, k is Boltzmann's constant of 8.617×10^{-5} eV/K, and T is temperature in degrees Celsius.

Calculations using Eq. (3) are shown in Fig. 2.10. All of the data are consistent with Eq. (3) except that of Ref. 47, which may have suffered some sort of experimental error. (One possibility for the differences in Seidel's etch rates is that the European source of KOH has a different amount of impurities, as discussed with regard to Fig. 2.8. Another possibility is that the reported temperatures are too high in Seidel's case (say, by 3°C), especially at high KOH concentrations, or they are too low in both Kendall and de Guel and Clark and Edell's work. We have favored the latter two sets of data because of the good agreement between them. We have observed errors as large as 10°C at Honeywell Microswitch when the magnetic stirrer was turned off and the controlling thermocouple was placed in the top of the solution and the wafers were etched in the bottom of a deep (13-cm solution) and wide (15×30 cm) container heated on a hot plate.[50] Our attempts to reproduce these results in a tall 2-liter beaker on a hot plate (on steadily, no temperature control) have resulted in only about a 3°C temperature difference between top (13 cm above the bottom) and bottom of the solution. In any case, this gives a wide possible range for temperature errors in an inadequately stirred or incorrectly temperature-controlled container containing KOH and water.)

There is no direct physics in either Eq. (1) for Si or Eq. (3) for SiO_2, except that the preexponential drops to zero when W goes to zero (pure water) and increases with W in a nonlinear fashion. The increase of the preexponential term with W for the SiO_2 etch rate, as well as for the Si etch rate of Eqs. (1a) and (1b), is due to the steady increase in the OH^- ion, which changes approximately with the second power of W, as shown in Eq. (2) and as carefully demonstrated by Glembocki et al.[49] Similarly, the activation energy, which is approximately linear in behavior for the etching of Si in Eqs. (1a) and (1b), increases slowly at low W and rapidly at high W for SiO_2 etching in KOH:water.

We do not attribute any special significance to the 2.5 power that shows up in the various equations except that it is close to the 2.0 power expected in the Glembocki analysis for the increase of the hydroxide ion concentration, and it is needed to fit the experimental results when approached in this empirical manner. The simplest way to think of the increasing preexponential is to consider the increase in the "attempt frequency" of the free OH^- ions as the KOH concentration increases. Of course, the empirically established increasing activation energy [from 0.795 at pure water to 0.96 eV at 60 wt% KOH of (Eq. 3)] ultimately causes the etch rate of SiO_2 [as well as that of Si in Eqs. (1a) and (1b)] to decrease at high KOH at any temperature. Physically, the decline in etch rate at high KOH is almost certainly due to the increasing hydration of the OH^- from Glembocki's model.

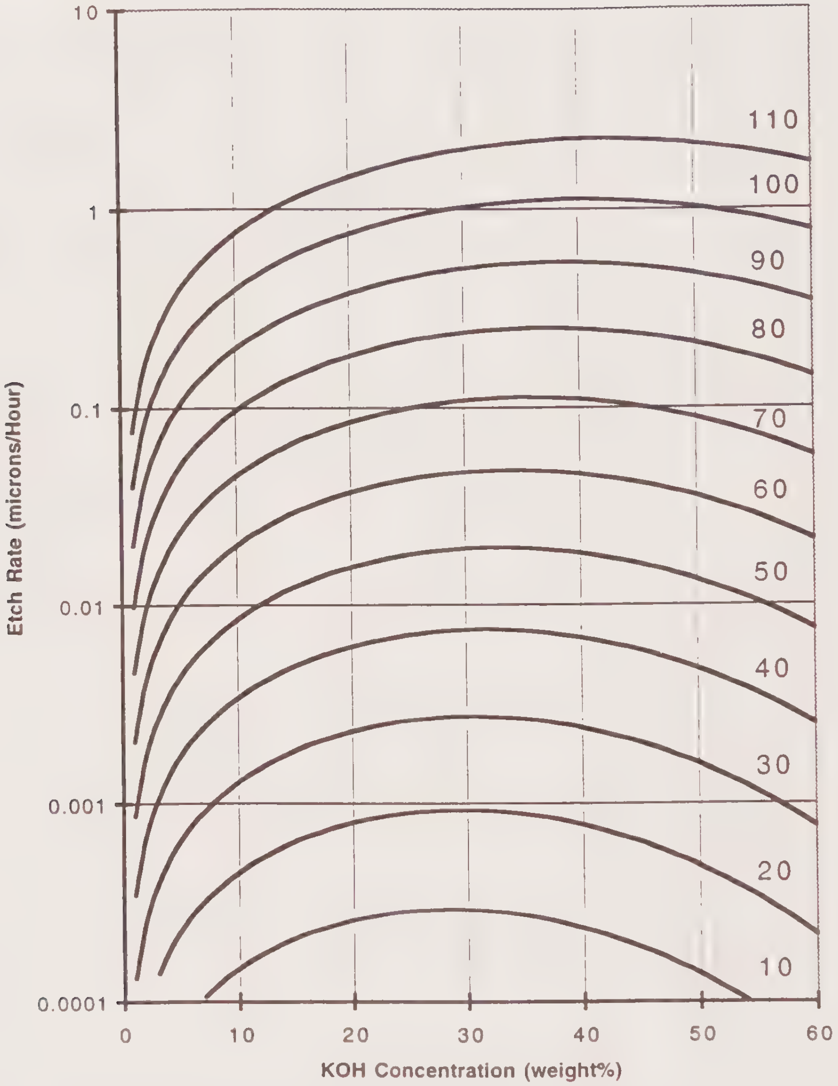


Fig. 2.10. Empirical etch rates of thermal SiO_2 in KOH:water.

Finally, we note that the addition of small amounts of ethylenediamine tetraacetic acid (EDTA), in particular the Na-free version of this chemical, has a surprisingly large effect on the etching of SiO_2 in KOH:water. For example, only 6 ppm by weight of EDTA nearly doubles the SiO_2 etch rate at 40°C in this solution.[50] (The increase is much less at 85°C.) This surprisingly large increase is thought to be the result of the EDTA somehow affecting the boundary layer thickness or steric properties of the interface. At the same time, EDTA has almost no measurable effect on the Si (100) etch rate. On the basis of the large decrease of the etch rate on the {110} by EDA[43] and other organics such as the alcohols, we would expect that the {110} etch rate would be decreased by EDTA and that the {111} etch rate would be increased. However, we have not measured these rates carefully.

2.7 FIVE WET ETCHING CHALLENGES

We have identified a group of fundamental questions based on the wet etching mechanisms and processes discussed in this chapter. They would be the questions we would study if we had the time and the support. However, since we are at present involved in other research, we would welcome other workers in these and related studies. The concepts are probably not all original, but they are interesting and stimulating. We recognize that different workers in the field of micromachining will have different ideas as to what are important challenges. For example, Richard Davis of Honeywell Microswitch wonders why a high etch rate on the {111} is often correlated with a low etch rate of SiO_2. He also would like to find some additives for the organic hydroxides (such as EDA and TMAH) that would reduce the etch rate on the {111} while still maintaining a low rate for the SiO_2. Davis would further like to find some wet-chemical etched structure that would provide economical through-wafer contacts without resorting to plasma etching. Honeywell Microswitch has already accomplished this on their pH sensor using (100) Si, but the wafer real estate used is larger than desired. It would be more efficient in terms of area use to etch very narrow slots from the front to back of the wafer on (110) wafers,[45] but the ion-sensitive field effect transistor (ISFET) would require significant redesign. The use of Lehmann's straight pores[17] on (100) wafers is also a candidate for this type of application, but such schemes have not yet been developed for these special pores.

2.7.1 Unidentified Zones in the HNA And Sirtl Ternaries

When the HNA ternary of Fig. 2.1 is inspected, it is apparent that there are a number of regions that could still be explored for special etching effects. The etch rate drops rapidly as the acetic corner is approached, but it is this region that includes several interesting etchants. For example, the large region to the right of the concentration-dependent etch and the extreme lower left (mostly acetic acid) are relatively unstudied. The addition of various metals and organics to the basic HNA system could also be useful for special purposes.

The so-called Sirtl ternary (Fig. 2.5) introduced in this work is also interesting in this regard. There are large regions that have not been studied in any systematic way. For example, the region between the dislocation pit producers and the mound reducers might result in an ambiguous etch that produces a polishing etchant, or one resulting in alternating mounds and pits on an atomic basis for texturing. Another region of interest would be along the dotted line in the lower right, which represents a rough estimate of the solubility limit of CrO_3 in the

HF:water solvent. The surface properties in either of these two regions may be exquisitely sensitive to defects or even to different crystal orientations, as will be discussed in the next section in some detail.

An example of splitting the difference between the pit-forming and the mound-forming etchants for the (1 1 1) of Fig. 2.4 would be the PM1 etchant, or "pit-mound 1" etchant. In this case, the 1 stands for the last number of the Miller index when the first two indices are 1 and 1. Similarly, the (0 0 1) could be denoted by (1 1 infinity) and hence by PM-I, where the I denotes infinity. The (1 1 0) equivalent of such an etchant would be called PM0, where we have assumed for simplicity that the mound-forming mixture is approximately the same for all orientations. Specifically, the PM1 mixture is located at 18, 52, and 30% of HF, water, and CrO_3 on the artificial volume ternary of Fig. 2.5. We say this is artificial because the CrO_3 is a dry chemical and we have chosen to express its "volume" percentage as a percentage of its molality relative to 20 *M*. This is a convenient number since it slightly exceeds the molal solubility of CrO_3 in water, which is 16.5 *M* at room temperature[82] (and since the molecular weight is 100, this is a solution of 1650 added to 1000 ml of water). Then the PM1 mixture would consist of 1000 ml of HF, 2900 ml of water, and 600 g (i.e., 30% of 2000 g) of CrO_3 powder, or in the nomenclature of the Sirtl family discussed earlier, 1:2.9 parts of 6 *M* CrO_3. One will note that this artificial volume concept is easy to visualize and use, but it assumes for simplicity that the specific gravities of the CrO_3-water mixture and the HF solutions are effectively 1.0.

2.7.1.1 Etching for atomic step and Miller index delineation

The first important special purpose we have in mind for the above chemical etchants is the ability to uniformly distribute and delineate single and double (and other low numbers of) atomic steps on slightly misaligned low index planes. This is readily accomplished by annealing at temperatures over 1000°C in ultrahigh vacuum systems as shown in Fig. 2.3, but has not been successfully accomplished in wet chemical systems to our knowledge except for the delineation of large steps in the KOH:water system and probably many other systems. This smoothing out of surfaces into single- and double-layer steps is quite important to some of the applications suggested in the next section, as well as certain catalytic applications not discussed here.

To understand why the larger step heights are more commonly observed during wet etching, one must realize that the atomic-size steps on a slow etching plane generally etch laterally much faster than the larger steps. This is due primarily to the greater availability of reactants for step etching on a slowly etching plane, and therefore less diffusion limitation than occurs near a fast-etching large step. The small steps then bunch up on a few slower moving steps that are slowed by impurity condensation or geometrical variation (kinks, etc.). This is quite similar to the way that traffic stacks up behind slow cars on narrow highways. The occasional breaking loose of high-energy vehicles is similar to the few fast-etching small steps in the open region between the large steps.

Examples of the movement of the larger steps (10 to nearly 1000 atoms high) during KOH anisotropic etching of (1 1 1) Si have been published.[45,52,53] The challenge would be to develop etchants that distribute these steps into single- or double-height thicknesses. This can probably be done using polishing etchants like PLE, CP4A, or SSE of Table 2.1, but at considerable sacrifice in surface flatness due to the diffusion limitations discussed earlier. The pH-modified BOE discussed earlier may be a reasonable starting point for producing such small steps on {111} surfaces,[27] but would probably not do a good job on other crystal planes (see Fig. 2.3 for the desired result).

The second important special purpose for one of the yet-unstudied regions of either the HNA or Sirtl ternaries (or other etching systems) would be to show particular Miller index surfaces with sharp boundaries. The best example of such a challenge might be to sharply delineate surfaces such as the (5 5 12) from the closely neighboring (2 2 5) and (3 3 7) surfaces at -1.0° and 0.7° on each side of the (5 5 12). These surfaces are not as unusual as one might think since they are located between the (1 1 2) and the (1 1 3) planes, as can be seen by dividing each of the indices by 2, 3, or 5. For example, we could say that the (5 5 12) is also the "(1 1 2.4)" surface that is located at 30.5° from the (0 0 1) plane.

We have chosen the above unusual planes because Baski et al.[37,38] have reported that the (5 5 12) is an almost perfectly flat atomic plane consisting of alternating (2 2 5) regions and two neighboring (3 3 7) planes. Furthermore, the (5 5 12) surface is evidently "flat" because the tensile stresses of the (2 2 5) are almost perfectly compensated by the compressive stresses of the two juxtaposed (3 3 7) plane regions. This appears to be an interesting set of planes to be delineated by some special etchant that is sensitive to small differences in stress (and local surface energies). For example, one might choose to briefly etch a slightly curved chemically polished surface that includes the relevant planes, or even a large Si sphere similar to those etched by Weirauch[83] and Hesketh and co-workers.[84] The etching should be brief and the spheres or the radii of curvature should be large to avoid the fast-etching planes that cause facets to form on such spheres, which would obliterate the information on closely spaced crystal planes.

For studies of crystal planes in the vicinity of the (0 0 1), one might prepare spherical or long cylindrical depressions using KOH:water at high concentration.

These mixtures have been shown to produce surprisingly smooth depressions in (0 0 1) surfaces, which should be useful in etching studies of this sort. Arrays of relatively large spherical depressions (micromirrors) can be formed quite easily by thinning wafers with preetched small pits into smooth depressions several millimeters in diameter, which have rather large radii of curvature of several centimeters.[85]

The challenge then is to demonstrate that different crystal orientations can be sharply delineated with a chemical etchant, without having to resort to expensive high-temperature heat treatments under ultrahigh vacuum conditions. Either the HNA or the Sirtl ternaries may point the way for such studies, perhaps with additives. There are two papers on the Sirtl system by the same authors[86] that should be quite helpful in understanding the electrochemistry of high vs. low HF/CrO_3 ratios (Sirtl vs. Seiter mixtures). They do not seem to be aware of the Seiter mound-forming properties, but their explanations appear to allow this possibility.

2.7.2 How Do Pits Relate To Interface States on Si?

In Section 2.2.5 on defect delineation etches, we mentioned the fact that the ideal Sirtl-type etchants for dislocation pits had a surprising linear relationship to the interface densities on the principal planes of Si. This linear relationship is accurate to within about 10%, which makes one believe that it might be meaningful. We grant that the surface energy should also follow the interface densities, but it is difficult to make any mechanistic connection to the pit-forming tendency of the three Sirtl etchants. We will leave this as a challenge for the reader.

2.7.3 Major Questions from Hydroxide Etching Mechanisms

2.7.3.1 Why do organic bases etch {100} faster than {110}?

The first challenge of this section is to develop a credible mechanism to explain the above question. We remind the reader that the etch ratio for the three low-index planes using KOH:water at high concentrations is 200:400:1 for the (100):(110):(111), respectively. If alcohol or other organics are added to the KOH solutions, the tendency is for the rate on the (110) to decrease, the (100) to stay about the same, and the (111) rate to increase significantly.

An extreme case of this organic base effect is the relative etch rates of ethylenediamine, or its cousin EDP. The three major planes then etch at approximately 200:100:10. This means that the (110) rate has dropped by a factor of 4, the (110) is relatively unaffected, and the (111) has increased by about an order of magnitude.[43] This is particularly surprising when it is realized that EDA is a strong base, not greatly unlike KOH:water. Nevertheless, the part of the EDA molecule that remains after creating the OH^- ions evidently interacts with the Si surfaces in an unusual fashion. It appears that the (110) is somehow strongly passivated compared with the KOH:water case, and the (111) surface is "unpassivated" to a great degree. This is just a working idea and the challenge would be to find other mixtures that would behave in a predictable manner in this or another model.

2.7.3.2 Can the "hydroxonium" mechanism be proven?

The mechanism proposed here and in some detail by Glembocki et al.[49] is based on the blocking of the (111) by the threefold tetrahedra made up of the OH^- ion complexed with three water ions, which we call "hydroxonium," as discussed earlier. The other major crystal planes are also affected uniquely, but are not blocked nearly as well as is the threefold symmetric (111). The challenge then is to "prove" this model using multiple-pass total internal reflection, as done by Higashi et al.,[27] or some other analysis method that detects hydrogen and other bonds on the surfaces (probably, though not necessarily, while they are immersed in KOH:water or other basic solution).

2.7.3.3 Why is there a large trace metal effect of KOH on Si?

We discussed in some detail the large variability in the etch rate of the (100), as in Fig. 2.8. Two very recent articles shed considerable light on this subject.[87,89] It was discovered that the addition of pure metals to a 33 wt% KOH:water solution changed the etch rates on (111) Si by very large factors. The changes induced by the pure metals (not the metal chlorides[88]) were generally less than about 5% on the (100).[88] The change in what they called the "anisotropy ratio," which is the vertical etch rate on the (100) divided by the lateral undercutting under the silicon nitride mask, was from about 60 for relatively pure KOH to as much as 130 when 500 ppm of Zn was added to the solution. The elements Al, Cr, Na, and Fe also increased the anisotropy ratio, in that order, but by lesser amounts than the Zn. Furthermore, none of these elements showed any sign of the saturation of the effect, so presumably even greater anisotropy ratios can be obtained by adding more than 500 ppm of Zn or Al. [The rates on the (110) were not measured].[88]

By contrast, the elements Cu and Ni both decreased this ratio by about 20% with the addition of only about 50 ppm and then the ratio became saturated for quantities up to 500 ppm. The saturation of the effect with Cu and Ni is almost certainly due to the limited solubility of these two metals inKOH:water, whereas the other metals apparently have much larger solubilities in this solution. We should also point out that the relative effect of the various metals is much stronger when the metals are introduced in quantities less than 20 ppm, thereby illlustrating the difficulty of exactly predicting the anisotropy ratios produced by different sources of KOH. Based on our personal experience, we do not believe that the European sources are any worse (or any better) than the American sources in this regard. A serious challenge to the various chemical companies would be to produce a good "semiconductor grade" of KOH that contained less than 1 ppm of all the metals mentioned here.

The second paper by Obermeier's group[89] treats the effect of large amounts of dissolved Si. The etch rates on the (100) are only modestly affected, even when 15 typical Si wafers of 100-cm diameter are dissolved in a liter of the 33 wt% KOH solution (our calculation based on their data). The biggest effect of the added Si is to increase the specific gravity from 1.33 to 1.56 with the addition of 190 g of Si to a liter of solution. There is also a considerable effect on the etching of convex corners by the Si with a shift from exposed {411} facets toward {311} facets and a roughening of the exposed facets.

The bigger challenge here is to explain the effect of the various metals and the much larger effect of adding Fe as the chloride mentioned earlier, which can easily double the etch rate on the (110) as shown in Fig. 2.7. This last observation again underlines the difficulty of predicting the precise etching properties of KOH solutions on Si for micromachining and other applications.

2.7.3.4 Determining the maximum anisotropy obtainable

Finally, we remind the reader that the (111) etch rate is not known with very good precision. It may be only 200 to 400 times as much as the rate on the {110} and it may be as slow as 1400 times slower than the (111) rate as measured by the incredible parallelism of the two walls of a narrow vertical stud.[90] As discussed earlier, there are several factors that make this very slow etch rate difficult to measure. This includes stress in the etch mask, oxide precipitates in the Si, mask alignment with respect to the {111}, and probably other factors. The ideal method of measurement may be the recent report of an optical interference method that bounces a laser beam off of the thin wafer while in the etching solution.[91] This method avoids the mask stress and alignment problems. Then if the thin wafer is precisely oriented with the {111} and the two sides are parallel, as well as properly heat treated as discussed earlier, a quite precise measurement could be taken using this optical method.

This is an important challenge and could allow many much higher anisotropy ratios to be developed, especially between the {110} and the {111}. This latter etch ratio is the critical one in obtaining higher aspect ratio grooves using wet chemical etching.

2.7.4 Narrow vertical grooves on the (100)?

In the perspective section, we discussed the possibility of fabricating CMOS circuits on the {110} to get more balanced CMOS due to the much higher hole mobility on this surface. Another advantage of the {110} is the ability to chemically etch very deep vertical grooves for making trench capacitors. A companion question for room temperature operation and higher is whether it will ever be possible to use wet chemistry to etch deep and narrow vertical grooves in the {100} for fabricating high-aspect ratio trench capacitors. We discussed earlier (Section 2.7.3.1) the strong effect that EDA had on the (110) and (111) etch rates, decreasing the first and increasing the second. If we could carry this to an extreme where the (110) etch rate is very strongly decreased and the (111) is strongly increased to the point where the (111) etches faster than the (110), then we could produce deep vertical grooves on the (100). This assumes that the (100) rate would now be markedly faster than either of the other two major low-index planes.

To produce the narrow grooves on the (100) under the above situation, one merely aligns the mask edges along the [011] direction as is done at present to produce V-grooves. However, now the {111} surfaces etch rapidly and leave exposed the perpendicular {110} on the walls of the now vertical grooves. The key issue is whether some aliphatic base (organic) etchant or some additive to KOH or one of its cousins can passivate the twofold symmetric {110} while at the same time destroying the passivating qualities on the {111} surfaces.[43] Some twofold long-chain molecule may accomplish both functions, but it will require some serious thinking to accomplish such a trick on Si wafers. Nevertheless, the discovery of such a wet etchant would be significant, since it would make a number of three-dimensional ICs and deep trench capacitors possible on a surface with low surface state density.

We should mention that narrow vertical grooves on InP surfaces have already been produced on {100} InP wafers because the mask can be aligned so that the fast-etching {111B} planes are slower than the slow-etching {110} planes in HCl solutions.[92] The challenge is to do the same on Si.

2.7.5 Why the Large EDTA Effect of KOH Etching of SiO_2?

The last challenge of this group of wet chemistry questions is to understand the factor of two enhancement in the etch rate of (100) Si when only a few parts of million of EDTA are added to KOH:water solutions, as discussed earlier (Section 2.6). Is it a steric effect, a boundary layer decrease, some chelating property of the EDTA on the SiO_2 surface, an effect on the water hydration, or some effect on the impurities in the KOH? Understanding this effect may help us to better understand the Si etching process, since Si on the (100) and (110) and SiO_2 have similar etch rate behaviors at different KOH concentrations.

2.8 FIVE NEW MICROMACHINING CHALLENGES

We have also identified five new structural concepts based on micromachining of Si that have not been studied so far, at least as far as we can determine. These are outside the plethora of applications discussed at recent conferences like “Transducers ‘97” in Chicago in June 1997.

2.8.1 Ordered Bubble Flows As Analogs for Superconductivity

Imagine dissolving {110} Si in KOH and producing deep vertical grooves separated by very thin vertical walls. In fact, imagine that the walls are so thin that the H_2 bubbles released during the etching process [see Reaction (2)] actually distort the thin walls. One can envisage an ordered array of bubbles in the grooves, a top view of which might look like the ones shown schematically in Fig. 2.11. We have seen such an effect while etching high-value trench-type capacitors. In that case the grooves were about 20 μm wide and 350 μm deep, and the walls were about 5 μm wide, as shown in Fig. 11 of Ref. 45. The bubbles (of about 25 μm diameter as viewed from the top) in neighboring grooves tended to avoid each other due to the geometrical hindrance of the bent walls. In addition, there was another row of bubbles in the alternate rows displaced along the length of the grooves as shown in Fig. 2.11, though they were more separated than in this figure.

The important question is: "Are there any favored bubble sizes and wall thicknesses and other geometrical constraints that allow the bubbles to flow more freely along their groove lengths?" An experimental and theoretical analysis of electrophoretic or gravity-induced bubble flow might develop into a powerful classical analog of several important phenomena in physics. For example, the strikingly small effective mass of the electron in materials like GaAs might be modeled by a squeezing (peristaltic) force of the phonons to help propel the electrons along the ordered lattice. Even more appropriate might be the ordered motion of electrons (or holes, or excitons) in a superconducting crystal lattice, where the lattice conspires to give easy passage to the charge carriers by peristaltic forces that might be somewhat analogous to the superbubble array shown here. We believe this is a non-trivial question that could open up some simple ways to look at physical concepts that are at present difficult to visualize.

From an experimental viewpoint, it also raises questions as to how to form bubbles in such structures without continuously dissolving the walls, for example, as shown in a recent paper using electrolytically generated gas bubbles.[93] An unusual type of liquid pump might also be constructed, with the liquid flow opposite to the bubble flow. This is reminiscent of electron flow in the valence band when holes flow in semiconductors.

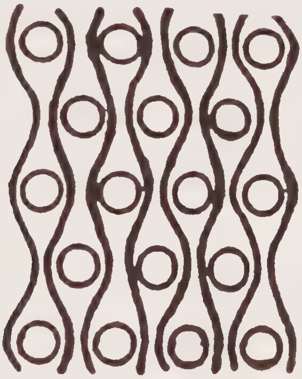


Fig. 2.11. Schematic view of hydrogen bubbles produced when etching deep grooves with deformable thin walls.

2.8.2 One-Way Flow in Porous Si Using Tethered Buckyballs

We have recently attempted to pass a vapor through a porous Si membrane in one direction only.[94] The idea is to make very small pores in Si of about 1 to 2 nm in diameter,[95] and then to plug them up on one side of the membrane with a tethered buckyball[96] (about 1 nm diameter), as shown schematically in Fig. 2.12. Unfortunately, our Russian-chemist source of tethered buckyballs returned to Europe so the experiment must now take other forms. "Gas rectifiers" of this type would enlarge our capabilities for making gas separations much more efficient, as well as for chemical gating. Methods to make the tethered buckyballs switchable would be the next challenge.

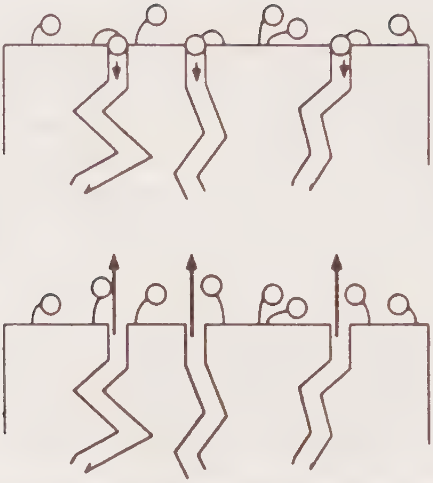


Fig. 2.12. Schematic of tethered buckyballs as one-way gas valves on one side of a porous Si membrane.

2.8.3 Deuterium Isotope Separation on {111} Si

The atomic-geometric, or steric effects on the wet etching of Si in KOH and other hydroxide etchants are striking indeed. In this and previous work, we have emphasized the role of the OH ion surrounded by three water molecules (which we call the "hydroxonium ion") as the dominant difference in the etch rates, which can vary as much as several hundred on different crystal planes. The question arises as to whether these large differences can somehow be used to separate deuterium from hydrogen. Based on this model, the incredibly slow etch rates on the {111} would appear to be extra sensitive to the bonding properties of the hydroxonium ion attached to the {111} atoms. Thus, we suggest that the D atoms (and even more so the tritium or T atoms) will interchange with their neighbors more slowly than the normal H atoms. This, as well as the strongly modified vibrational properties, is quite likely to make the chemical bonding much stronger (or much weaker) for the D atoms attached to the {111} surface.

The implications of the above comments are manifold. First, the etching properties of KOD:D_2O on {111} Si may be quite different from those of KOH:H_2O. Second, the ratio of D to H in a "normal" KOH:H_2O solution surrounding a high area of {111} wafers may be quite different from the truly normal ratio of about 1:6000.[97] The latter suggests that if the D atoms are more strongly attracted to the {111} surfaces, then the surrounding water will be "supernormal" (impoverished of D atoms). On the other hand, if the D atoms are repelled from the {111} surfaces, then the water will be enriched in D atoms. A simple calculation shows that the number of {111} atomic sites (on a multiple vertical groove wafer of {110} Si) can be made similar to the number of D atoms in a small volume of water that is appropriately immersed in nitrogen or another gas to reduce the interchange of D and H with the water in the atmosphere.

Whether impoverished or enriched, useful separations can be performed using the above concepts. If the surrounding water is greatly impoverished of D, then supernormal water can be harvested. If it is greatly enriched in D, heavy water can be obtained. (Porous Si can have a much higher surface area of {111} Si and can also be used for these experiments, but may require more time to refresh than the vertical grooves.) The challenge of these concepts is to perform the above and related experiments and perhaps find optical, thermal, electrical, or other ways to harvest or rapidly refresh the micromachined structures for continued use.

2.8.4 Vertical Semipermeable Membrane Structures

We have fabricated[98] interdigitated grooves separated by vertical semipermeable membranes made using porous Si, as shown in Fig. 2.11 and first shown in a scanning electron microscope photo of Fig. 4 of Ref. 53. Figure 2.13 shows a microreaction chamber that allows a chemical reaction to occur in a tightly constrained volume after the reactants A and B each enter in a set of interdigitated grooves and diffuse through the membranes into volume C. This follows a general reaction of the form:

$$\text{CtA and CtB}$$

$$A + B \rightarrow A^* + B^* \rightarrow C^* + BP \qquad (4)$$

where A and B are the unexcited reactants, A^* and B^* are the momentarily excited reactants after passing through the separate membranes (each with a different catalyst CtA and CtB if desired), C^* is the desired product with perhaps a short excited lifetime, and BP stands for "by-products." The possible short excited lifetime in a very small volume of the product means that highly explosive reactions can be undertaken with little danger. A less complicated system that has a similar advantage has recently been reported.[99] In the latter, the author demonstrated the nonexplosive reaction of NH_3 and O_2, as well as 20%H_2 and O_2, in their microreaction chambers.

An important example mentioned in the literature[100] is when A is methane, B might be H_2SO_4 vapor, C^* is methyl bisulfate, and the by-products would be water and sulfur dioxide. The methyl bisulfate, which is a protected form of methanol, is then added to water to form methanol while regenerating H_2SO_4, which is returned to the original process to repeat the reaction. This important "methane to methanol" reaction could convert the natural gas waste product of

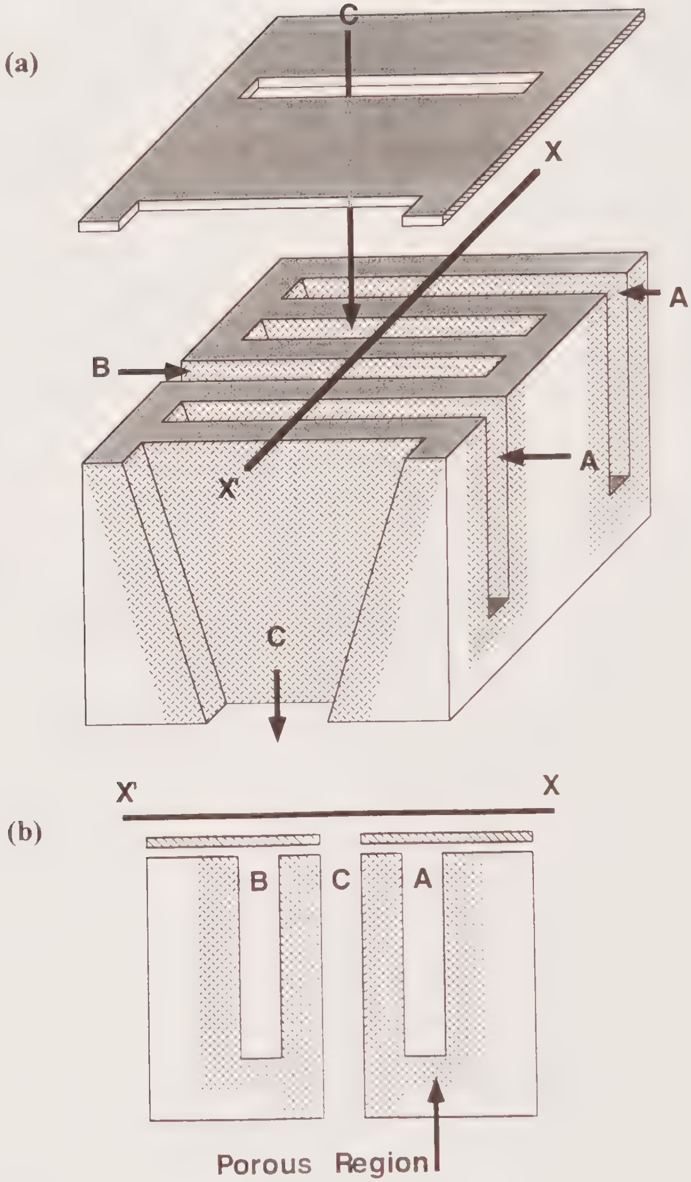


Fig. 2.13. Perspective view (a) and cross-sectional view (b) across XX' of a micromachined chemical reaction chamber showing interdigitated deep grooves, semipermeable vertical membrane made with porous Si, and grooved lid to allow flushing of products and to separate neighboring wafers in a stack.

oil wells to an easily transportable liquid.[100] It occurs at a mild temperature of 180°C, which is well within the range of Si substrates. This particular reaction used Hg as a catalyst, but we would propose using a more environmentally friendly pair of catalysts such as Pt and Pd. For example, these two catalysts could be introduced using dilute HF contaminated with a few parts per million of each metal into separate interdigitated grooves during manufacture of the structure. (Note the different thicknesses of the vertical studs in the cross-sectional view of Fig. 2.13. This prevents intermixing of the catalysts in the membranes.)

It should also be noted that the gases A and B must somehow be kept from mixing before they enter the two sides of the structure of Fig. 2.13. This might be done by stacking the structures by the hundreds or even thousands and press fitting the square stacks of chips into a circular pipe. Then gases A and B could enter on two edges of the pipe, and the products and by-products could be forced to exit down the group of interconnected C-holes down the axis of the "pipe factory." A special connector at the two ends of the pipe could accomplish this complex task.

We believe that such structures will be built in the near future, and that their heart will be the vertical semipermeable membranes. Simpler structures would involve electrophoresis down such grooves, with controlled leaks into neighboring chambers using the porous walls. Our difficulty with thinking in three dimensions is probably our biggest challenge in regard to inventing such structures, but the next generation of micromachinists will be much better at it.

2.8.5 Directional-Facing Atomic Steps for Chiral Boxes

In 1990 a method was suggested for producing square holes in {100} Si with clockwise(CW)-facing atomic steps on four walls.[46,53] This turns out to be quite difficult using anisotropic etchants such as KOH since the obvious approach of using a square etch mask opening that is misaligned from the perfect {111} intersections will simply not work. This is because the misaligned opening leads to a perfectly faceted inverted pyramidal opening that is rotated with respect to the original opening, but that contains no misalignment steps. However, truly misaligned holes such as those shown in Fig. 2.14 can be produced using laser-enhanced etching of {100} Si or other orientations, or even other materials. For example, the laser beam can have a square cross section, which can be slightly misaligned from the major low-index planes during the laser etching process.[53]

What would be the uses of such misalignment steps? Forgetting square holes for a moment, such atomic steps can form the basis for producing quantum wires by obliquely evaporating or otherwise depositing a conducting material onto the steps.[101] These are important applications and can "become useful for the study of two- versus one-dimensional magnetism, for magnetoresistive films, and in the design of anisotropic materials."[101] However, our question is a bit more complex, and asks how such CW-facing steps might be used if they were present in a square hole.

One application for the unusual square holes might be to separate right-handed molecules from left-handed molecules that are completely identical except for their "handedness," or their "chirality." One can imagine that such molecules dissolved in an appropriate solvent that was passed through a large set of holes like that shown in Fig. 2.14 might favor the passage of right- (or left) handed molecules. This would be due to the reflections of the molecules with the CW-facing atomic steps in the holes of the "chiral filter." This is a very important commercial application since the chemical activity of a right-handed molecule can be radically different from that of an otherwise identical left-handed molecule.[102] There are many biological and other reactions that could greatly benefit from a universal chiral filter.[57]

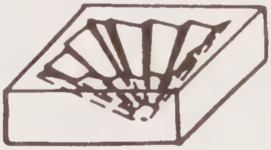


Fig. 2.14. Misaligned square hole on {100} Si for chiral molecule separations and other applications.

A more speculative concept would use such hole structures to create a CW molecular swirl of gas or liquid molecules, namely, a large set of micro-tornadoes, all operating at once.[46,53] If properly activated by current or light pulses on the atomic steps, this could conceivably cause a pressure gradient that would give the structure aerodynamic lift (a flying carpet effect). If structures molded from such templates were placed in airplane wings, then it was suggested that sudden wind shears could be compensated for by pulsed current down the atomic steps. This would create lift by an electronic feedback process based on the local wind forces on the wing.[46]

Are such chiral box structures really feasible and useful? We are not sure, but we see no reason they cannot be fabricated and tested for some of these unusual but highly leveraged applications.

2.9 CONCLUSION

The ten challenges described here were all derived from our studies and cogitations on bulk micromachining and wet chemical etching over the past 28 years or so. There are more that we haven't mentioned, some of which we are working on. In any case, we hope we have stimulated the readers to think in other realms. We also believe that some, though probably not all, of these challenges will be "old hat" to the next generations of micromachiners. In fact, most of the wildest imaginings of today will likely be obvious to our grandchildren, and they will wonder why we missed some of the simple ones.

ACKNOWLEDGMENTS

The authors appreciate the advice and encouragement of colleagues at Honeywell Microswitch in Richardson, Texas, at Sandia National Laboratories in Albuquerque, New Mexico; and at the University of New Mexico in Albuquerque. We also thank Thomas Digges at Virginia Semiconductor Inc., Mark Larson at MOXTEK Inc., and Alex Hoelke at the University of Cincinnati for permission to quote unpublished data.

REFERENCES

1. M. Aoki, K. Yano, T. Masuhara, and K. Shimohigashi, “Fully symmetric cooled CMOS on (110) plane,” *Trans. IEEE Elect. Dev.* **36**, 1429 (1989).

2. T. J. Brosnihan, J. M. Bustillo, A. P. Pisano, and R. T. Howe, in *Proceedings Transducers 97*, Int'l Conf. on Solid State Sensors and Actuators, Abstr. 2D2.04, 1997; W. H. Juan and S. W. Pang, Abstr. 1A4.09P.

3. B. Schwartz and H. Robbins, “Chemical etching of silicon: IV. Etching technology,” *J. Electrochem. Soc.* **123**, 1903 (1976).

4. W. R. Runyan and K. E. Bean, *Semiconductor Integrated Circuit Processing Technology*, Addison-Wesley, Reading, MA, (1990).

5. W. Kern and C. A. Deckert, in *Thin Film Processes*, John L. Vossen and W. Kern, eds., Chap. V-I, Academic Press, New York, (1978).

6. H. Muraoka, T. Ohhashi, and Y. Sumimoto, in *Semiconductor Silicon 1973*, H. R. Huff and R. R. Burgess, eds., pp. 327–338, Electrochemical Society, Pennington, NJ, (1973).

7. M. Barrett, M. Dennis, D. Tiffin, Y. Li, and C. K. Smith, “2-D dopant profiling in VLSI devices using dopant selective etching,” *IEEE Electron Dev. Lett.* **16**, 118 (1995).

8. A. Montiel and D. L. Kendall, “Dielectric isolation using vertical etching (DIVE),” extended abstracts, Abstr. 180, Electrochemical Society Meeting, (1978).

9. S. M. Hu and D. R. Kerr, "Observation of etching of n-type silicon in aqueous HF solutions," *J. Electrochem.* **114**, 414 (1967).

10. K. H. Beckman, "Investigation of the chemical properties of stain films on silicon by means of infrared spectroscopy," *Surface Sci.* **3**, 314 (1965).

11. D. L. Kendall and R. A. Davis, unpublished data, Honeywell Microswitch, Richardson, TX.

12. A. Hoelke, Univ. of Cincinnati, private communication.

13. G. Bomchil, A. Halimaoui, and R. Herino, "Porous silicon: The material and its applications to SOI technologies," *Microelectron. Eng.* **8**, 293 (1988).

14. Terry Guilinger, Sandia National Laboratories, private communication;. C. L. Levy-Clement, A. Lagoubi, and M. Tomkiewicz, "Morphology of porous n-type Si obtained by photoelectrochemical etching: Correlations with material and etching parameters," *J. Electrochem. Soc.* **141**, 958 (1994).

15. P. Maccagnani, R. Angelucci, P. Pozzi, A. Poggi, L. Dori, G. C. Cardinali, and P. Negrini, "Thick oxidized silicon as thermo-insulating material for high temperature operating thin and thick film gas sensors," in Ref. 2, Abstr. 1D3.04.

16. J-H. Sim, S-H. Hahm, J-Hee Lee, J-Hyun Lee, I-S. Yu, J-Sup Kim, "Eight-beam piezoresistive accelerometer fabricated by using a selective porous silicon etching method," in Ref. 2, Abstr. 4B2.03.

17. S. Ottow, V. Lehmann, and H. Foll, "Processing of three dimensional microstructures using macroporous n-type Si," *J. Electrochem. Soc.* **143**, 385 (1996) and included references.

18. W. Kern and D. A. Puotinen, "An efficient cleaning sequence for silicon wafers," *RCA Review* **31**, 187 (1970).

19. D. J. Antilla, M. V. Tilli, M. Schaekers, and C. L. Claeys, "Effect of chemicals on metal contamination on silicon wafers," *J. Electrochem. Soc.* **139**, 1180 (1992).

20. A. Mayer and S. Shwartzman, "Megasonic cleaning: a new cleaning and drying system for use in semiconductor processing," *J. Elect. Mat.* **8**, 885 (1979).

21. See Ref. 5, p. 411.

22. Ref. 5, pp. 412–413.

23. T. Ohmi and T. Shibata, "Scientific ULSI manufacturing in the 21st century," *Electrochem. Soc. Interface,* **1**, 32, (1992).

24. See Ref. 23, p. 35; T. Ohmi, M. Nakamura, A. Ohki, K. Kawada, and K. Hirao, "Trace moisture analysis in specialty gases," *J. Electrochem. Soc.* **139**, 2654 (1992); T. Ohmi, "Total room temperature wet cleaning of silicon surfaces," *Semiconductor International*, p. 323 (1993).

25. K. Kochiek, "Measurement of water adsorption layers on metal surfaces," *Metrologia* **18**, 153 (1982).

26. R. Carpio and D. L. Kendall, unpublished data at INAOE, Puebla, Mexico, (1976).

27. G. S. Higashi, Y. J. Chabal, G. W. Trucks, and K. Raghavachari, "Ideal hydrogen termination of the Si(111) surface," *Appl. Phys. Lett.* **56**, 656 (1990).

28. D. J. Chadi, "Stabilities of single-layer and bilayer steps on Si(001) surfaces," *Phys. Rev. Lett.* **59**, 1691 (1987).

29. B. S. Swartzentruber, C. M. Matzke, D. L. Kendall, and J. E. Houston, "STM measurements of step-flow kinetics during atom removal by low-energy-ion bombardment of Si(001)," *Surf. Sci.* **329**, 83 (1995).

30. A. J. Hoeven, D. Dijkkamp, E. J. van Loenen, J. M. Lenssinck, and J. Dieleman, "Island and step structures on molecular beam epitaxy grown Si(001) surfaces," *J. Vac. Sci. Technol.* **A8**, 207 (1990).

31. W. C. Dash, "Copper precipitation on dislocations in silicon," *J. Appl. Phys.* **27**, 1193 (1956).

32. E. Sirtl and A. Adler, "Chromic oxide:HF:water solutions as a specific system for delineating dislocations on silicon," *Z. Metallk.* **52**, 529 (1961).

33. D. G. Schimmel, "Defect etch for <100> silicon evaluation," *J. Electrochem. Soc.* **126**, 479 (1979).

34. K. H. Yang, "An etch for the delineation of defects in silicon," Ref. 33, **131**, 1140 (1984).

35. H. Seiter, "An integrational etching solution for delineation of dislocations in Si," in *Semiconductor Silicon 1977*, H. R. Huff and E. Sirtl, eds., p. 87, Electrochemical Society, Pennington, NJ, (1977).

36. See Ref. 4, p. 74.

37. S. C. Irwin, A. A. Baski, and L. J. Whitman, "Structure and stability of Si(114)-(2x1)," *Phys. Rev. Lett.* **77**, 687 (1996).

38. A. A. Baski, S. C. Irwin, and L. J. Whitman, "A stable high-index surface of silicon: Si(5 5 12)," *Science* **269**, 1556 (1995).

39. D. L. Kendall, Bruce Shipley, Robert Whitley, Anita Advaney, Andrew Hecht, unpublished research.

40. J. F. Secco d'Aragona, "A delineation etchant for stacking faults in Si," *J. Electrochem. Soc.* **119**, 948 (1972)

41. M. Wright Jenkins, "A new preferential etch for defects in silicon crystals," *J. Electrochem. Soc.* **124**, 757 (1977).

42. W. K. Zwicker and S. K. Kurtz, "Anisotropic etching of silicon using electrochemical displacement reactions," Ref. 6, p. 315.

43. D. L. Kendall, K. J. Malloy, and C. B. Fleddermann, "Critical technologies for the micromachining of silicon," in *Semiconductors and Metals*, K. T. Faber and K. J. Malloy, eds., Vol. 37, pp. 293–337, Academic Press, New York (1992).

44. K. E. Bean, "Anisotropic etching of silicon," *IEEE Trans. Electron. Dev.*, **ED-25**, 1185 (1978).

45. D. L. Kendall, "Vertical etching of silicon at very high aspect ratios," *Ann. Rev. Mat. Sci.* **9**, 373 (1979).

46. D. L. Kendall and G. R. de Guel, "Orientations of the third kind: the coming of age of (110) silicon," in *Micromachining and Micropackaging of Transducers*, C. D. Fung et al., eds., pp. 107–124, Elsevier Science, Amsterdam (1985).

47. H. Seidel, L. Csepregi, A. Heuberger, and H. Baumgartel, "Anisotropic etching of crystalline silicon in alkaline solutions, I.," *J. Electrochem. Soc.* **137**, 3612 (1990).

48. L. D. Clark, Jr. and D. J. Edell, "KOH:H2O etching of (110) Si, SiO2, and Ta: an experimental study," in *Proc. IEEE Micro-Robots and Teleoperators Workshop*, (Cat.87TH0204-8) (1987).

49. P. Glembocki, G. R. de Guel, and D. L. Kendall, "Hydration model for anisotropic etching of Si in aqueous alkali hydroxides," *J. Electrochem. Soc.* **138**, 1055 (1991).

50. R. Davis and D. L. Kendall, unpublished data, Honeywell Microswitch.

51. P. M. M. C. Bressers, J. J. Kelly, J. G. E. Gardeniers, and M. Elwenspoek, "Surface morphology of p-type (100) silicon etched in aqueous alkaline solution," *J. Electrochem. Soc.* **143**, 1744 (1996).

52. D. L. Kendall, "On etching very narrow grooves in silicon," *Appl. Phys. Lett.* **26**, 195 (1975).

53. D. L. Kendall, "A new theory for the anisotropic etching of Si and some underdeveloped chemical micromachining concepts," J. *Vac. Sci. Technol.* **A8**, 3598 (1990).

54. P. E. Freeland, K. A. Jackson, C. W. Lowe, and J. R. Patel, "Precipitation of oxygen in silicon," *Appl. Phys. Lett.* **30**, 31 (1977).

55. Mark Larson, MOXTEK Inc.; Tom Digges,Jr., Virginia Semiconductor Inc.; Don L. Kendall, Univ. of New Mexico, unpublished data.

56. T. Iwasaki, Y. Tsumori, K. Nakai, H. Haga, K. Kojima, and T. Nakashizu, "Influence of cooling condition during crystal growth of CZ-Si on oxide breakdown property," *J. Electrochem. Soc.* **143**, 3383 (1996).

57. N. P. M. Tuck, W. F. Yager, B. de Lange, and B. L. Feringa, "Dynamic control and amplification of molecular chirality by circular polarized light," *Science*, **273**, 1686 (1996).

58. T. A. Kwa, P. J. French, R. F. Wolffenbuttel, P. M. Sarro, L. Hellemans, and J. Snauwaert, "Anisotropically etched Si mirrors for optical sensor applications," *J. Electrochem. Soc.* **142**, 1226 (1995).

59. E. A. Irene, E. Tierney, and J. Angilello, "A viscous flow model to explain the appearance of high density thermal oxidation at low oxidation temperatures," *J. Electrochem. Soc.*, **129**, 2594 (1982); L. M. Landsberger and W. A. Teller, "Two-step oxidation experiments to determine structural and thermal history

effects in thermally grown SiO2 films on Si," *J. Electrochem. Soc.*, **137**, 2825 (1990).

60. K. Ohwada, Y.Negoro, Y. Konoka, and T. Oguchi, "Uniform groove-depths in (110) Si anisotropic etching by ultrasonic waves and application to accelerometer fabrication," *Sensors and Actuators*, **A50**, 93 (1995).

61. J. G. Hooley, "The kinetics of the reaction of silica with group I hydroxides," *Can. J. Chem.* **39**, 1221 (1961).

62. S. M. Repinski and S. V. Pershkova, "Kinetics of the dissolution of silicon in aqueous sodium hydroxide," *Russ. J. Phys. Chem.* **46**, 519 (1972) *Russ. J. Phys. Chem.* 4S.

63. L. D. Clark, J. L. Lund, and D. J. Edell, "Cesium hydroxide [CsOH]: A useful etchant for micromachining silicon," *Tech. Digest*, IEEE Solid State Sensor and Actuator Workshop, pp. 5–8, (1989).

64. T. Sethi and P. J. Hesketh, Univ. of Illinois at Chicago, private communication; T. Wang, S. Surve, and P.J. Hesketh, "Anisotropic etching of silicon in rubidium hydroxide," *J. Electrochem. Soc.* **141**, 2493 (1994).

65. H. Seidel et al., "Anisotropic etching of silicon in alkaline solutions, II. Influence of dopants," Ref. 47, **137**, 3626 (1990).

66. E. Steinsland, M. Nese, A. Hanneborg, R. W. Bernstein, H. Sandmo, and Kittilsland, "Boron etch-stop in TMAH solutions," in *Proc. 8th Int. Conf. on Solid-State Sensors and Actuators*, Eurosensors IX, p. 190, (1995).

67. L. M. Landsberger, S. Nasan, M. Kahrizi, and M. Paranjape, "Conditions for the formation of hillocks during alkaline etching of silicon," *J. Microelectromech. Syst.* **5**, 106 (1996).

68. M. M. Abu-Zeid, D. L. Kendall, G. R. deGuel, and R. Galeazzi, "Effect of geometry on EDA:H2O etching of silicon," extended Abstr. 85-1, Electrochemical Society meeting, Abstr. 275, (1985).

69. A. Reismann, M. Berkenbilt, S. A. Chan, F. B. Kaufman, and D. C. Green, "The controlled etching of silicon in catalyzed ethylenediamine-pyrocatechol-water solutions," *J. Electrochem Soc.* **126**, 1406 (1979).

70. O. Tabata, R. Asahi, H. Funabashi, and S. Sugiyama, "Anisotropic etching of silicon in (CH3)4NOH solutions," *1991 IEEE MEMS Proc.*, pp. 811–814, (1991).

71. U. Schnakenberg, W. Benecke, and P. Lange, "TMAHW etchants for silicon micromachining," *1991 IEEE MEMS Proc.*, pp. 815–818, (1991).

72. U. Schnakenberg, W. Benecke, B. Loechel, S. Ullerich, and P. Lange, "NH4OH-based etchants for silicon micromachining: Influence of additives and stability of passivation layers," *Sensors and Actuators* **A24**, 1 (1991); U. Schnakenberg, W. Beneke, and B. Loechel, "NH4OH-based etchants for silicon micromachining," ibid **A21-A23**, 1031 (1990).

73. J. E. A. M. van der Meerakker and M. H. M. van der Steten, "A mechanistic study of silicon etching in NH3/H2O2 cleaning solutions," *J. Electrochem. Soc.* **137**, 1239 (1990).

74. D. B. Lee, "Anisotropic etching of silicon," *J. Appl. Phys.* **40**, 4569 (1969).

75. M. Mehregany and S. Senturia, "Anisotropic etching of silicon in hydrazine," *Sens.Act.* **13**, 375 (1988).

76. H. Linde and L. Austin, "Wet silicon etching with aqueous amine gallates," *J. Electrochem. Soc.* **139**, 1170 (1992).

77. H. G. Linde and L. W. Austin, "Heterocyclic catalysts for enhanced silicon oxidation and wet chemical etching," *Sens. Act. A*, **49**, 167 (1995).

78. R. M. Finne and D. L. Klein, "A water-amine-complexing agent system for etching silicon," *J. Electrochem. Soc.* **114**, 965 (1967).

79. M. Josse and D. L. Kendall, "A rectangular profile diffraction grating from single crystal Si," *Appl. Opt.* **19**, 72 (1980).

80. See Ref. 4, pp.264–267

81. D. J. Monk, D. S. Soane, and R. T. Howe, "Determination of the etching kinetics for the hydrofluoric acid/silicon dioxide system," *J. Electrochem. Soc.* **140**, 2339 (1993).

82. *Perry's Chemical Engineers' Handbook*, 6th Ed., p. 3–11, (1984).

83. D. F. Weirauch, "Correlation of the anisotropic etching of single crystal silicon spheres and wafers," *J. Appl. Phys*. **46**, 1478 (1975).

84. C. Ju and P. J. Hesketh, "High index plane selectivity of silicon anisotropic etching in aqueous potassium hydroxide and cesium hydroxide," *Thin Solid Films*, **225**, 58 (1992).

85. D. L. Kendall, W. P. Eaton, R. Manginell, and T. G. Digges, "Micromirror arrays using KOH:H2O micromachining of silicon for lens templates, geodesic lenses, and other applications," *Opt. Eng.* **33**, 3578 (1994).

86. J. E. A. M. van der Meerakker and J. H. C. van Vegchel, "Silicon etching in CrO3-HF solutions: I. High [HF]/[CrO3 ratios," *J. Electrochem. Soc*. **136**, 1949 (1989); ibid, "II. Low [HF]/CrO3 ratios," **136**, 1954 (1989).

87. A. Hein, O. Dorsch, and E. Obermeier, "Effects of metallic impurities on anisotropic etching of silicon in aqueous KOH solutions," Ref. 2, Abstr. 2D3.07P.

88. A. Hein and E. Obermeier, private communication.

89. O. Dorsch, A. Hein, and E. Obermeier, "Effect of silicon content of aqueous KOH on the etching behavior of convex corners in <100> single crystal silicon," Ref. 2, Abstr. 2D3.06.

90. S. Zaidi and S. Brueck, Univ. of New Mexico, private communication.

91. E. Steinsland, T. Finstad, A. Ferber, and A. Hanneberg, "In situ measurement of etch rate of single crystal silicon," Ref. 2, Abstr. 2D3.12P.

92. L. A. Coldren, K. Furuya, B. I. Miller, "On the formation of planar-etched facets in GaInAsP/InP double heterostructures," *J. Electrochem. Soc*. **130**, 1918 (1983).

93. W. Olthuis, A. Volanschi, and P. Bergveld, "Dynamic tension measured with an integrated sensor-actuator device using electrolytically generated gas bubbles," in *Proc. Transducers '97*, Abstr. 3D3.01, p. 1959, (1997).

94. R. A. Shoultz, T. Guilinger, and D. L. Kendall, unpublished research.

95. See Refs. 13 and 14 and included references.

96. F. Diederich and C. Thilgen, "Covalent fullerene chemistry," *Science* **271**, 317 (1996).

97. *CRC Handbook*, 64th Ed, 1983–1984, p. B-17, CRC, Boca Raton, FL.

98. R. Manginell, D. L. Kendall, and T. Guilinger, unpublished research.

99. R. Srinivasan, S. L. Firebaugh, I-M Hsing, J. Ryley, M. P. Harold, K. F. Jensen, and M. A. Schmidt, "Chemical performance and high temperature characterization of micromachined chemical reactors," Ref. 2, Abstr. 1C3.01.

100. R. A. Periana, "A mercury-catalyzed high yield system for the oxidation of methane to methanol," *Science* **259**, 340 (1993).

101. T. Jung, R. Schittler, J. K. Gimzewski, and F. J. Himpsel, "One-dimensional metal structures at decorated steps," *Appl. Phys. A*, **61**, 467 (1995) and included references.

102. M. Yuzaki, D. Forrest, T. Curran, and J. A. Connor, "Selective activation of calcium permeability by aspartate in purkinje cells," *Science* **273**, 1112 (1996).

CHAPTER 3

Applications of Dry Etching to Microsensors, Field Emitters, and Optical Devices

Stella W. Pang
University of Michigan
Department of Electrical Engineering and Computer Science

CONTENTS

3.1 Introduction / 100

3.2 Dry Etching of Polymers / 101
    3.2.1 Etch characteristics for thick polymers / 102
    3.2.2 Electroplating Ni through polymers / 105
    3.2.3 Micromolding of ceramics / 108

3.3 High Aspect Ratio Si Microstructures / 110
    3.3.1 Etching of high aspect ratio Si microstructures / 111
    3.3.2 Deep etch-shallow diffusion technology / 116
    3.3.3 Laterally driven comb resonators in Si / 120

3.4 Si Field Emitters with Sharp Tips / 122
    3.4.1 Control of emitter profile by dry etching / 123
    3.4.2 Sharpening field emitter tips / 127
    3.4.3 High density sharp Si field emitter arrays / 131

3.5 Fabrication of Mirrors and Lasers by Dry Etching / 134
    3.5.1 Si mirrors and modulators / 135
    3.5.2 GaAs mirrors and waveguides / 138
    3.5.3 Horizontal distributed Bragg reflector mirrors / 142

3.6 Summary / 147
    References / 149

3.1 INTRODUCTION

Precise micromachining technologies for high performance microsensors, field emitters, and optical devices have been developed. A number of requirements have to be met for these devices including control in etch rate, profile, selectivity, surface morphology, and etch-induced damage. As the density, sensitivity, and/or speed of these devices increase, high aspect ratio microstructures with narrow width and large depth are needed. Therefore, improvements in dry etching technologies are necessary to meet the multiple requirements of these advanced devices.

In conventional reactive ion etching (RIE), the reactive species in the plasma are generated with an rf power supply connected to the substrate stage. In order to maintain a reasonable etch rate and vertical profile, high rf powers are typically used. Under this situation, the samples to be etched are bombarded with high energy ions, which could result in substantial degradation in the device characteristics. The etch selectivity is also low due to the high energy ions. High pressure, typically above 10 mTorr, is also needed in RIE in order to maintain a stable plasma. This could result in additional surface roughness and an undercut profile. Therefore, conventional RIE is not suitable for fabricating high performance devices with high aspect ratio features since it does not provide independent control of ion energy and ion flux. This problem can be resolved by using a high density source to generate the plasma and rf power on the stage to control the ion energy.

To improve the dry etching technology, a plasma system with an electron cyclotron resonance (ECR) source and an rf-powered electrode is used as shown in Fig. 3.1. Microwave power at 2.45 GHz is coupled into the ECR cavity through a microwave input probe. In the presence of a magnetic field, the electrons resonate in a circular motion with a resonant frequency qB/m_e, where q is the charge of an electron, B is the static magnetic field strength, and m_e is the mass of an electron. The magnetic field strength B corresponding to the ECR condition is 875 G. This results in a very efficient absorption of microwave energy by the resonating electrons. The electrons are accelerated in a cyclotron motion by the microwave energy and they ionize or dissociate the gases in the plasma system.

The high density plasma generated with the ECR source results in much faster etch rates and the low pressure provides more vertical etch profiles compared to those obtained from the conventional RIE system. An rf power supply at 13.56 MHz is connected to the substrate stage for independent control of the ion energy. The rf power can be kept low without loss of throughput or plasma stability since the plasma density is mainly controlled by the microwave power. This results in significant reduction of the ion energy in a plasma generated with the ECR source, and very low damage and high selectivity dry etching can be achieved.

In this chapter, the micromachining technologies used to fabricate microsensors, field emitters, and optical devices are reported. By providing precise control of the etch characteristics, high aspect ratio microstructures that can provide substantial improvements in device performance are demonstrated.

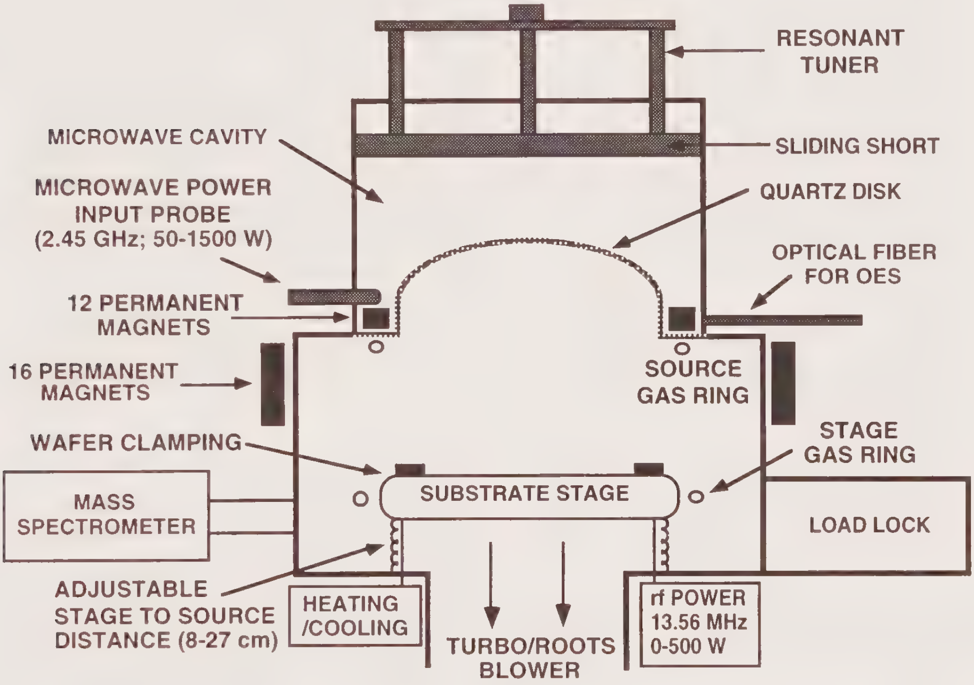

RESONANT TUNER
MICROWAVE CAVITY
SLIDING SHORT
MICROWAVE POWER INPUT PROBE (2.45 GHz; 50-1500 W)
QUARTZ DISK
OPTICAL FIBER FOR OES
12 PERMANENT MAGNETS
16 PERMANENT MAGNETS
SOURCE GAS RING
WAFER CLAMPING
STAGE GAS RING
MASS SPECTROMETER
SUBSTRATE STAGE
LOAD LOCK
ADJUSTABLE STAGE TO SOURCE DISTANCE (8-27 cm)
HEATING /COOLING
TURBO/ROOTS BLOWER
rf POWER 13.56 MHz 0-500 W

FIG. 3.1 Schematic of plasma system with an ECR source and an rf-powered electrode.

3.2 DRY ETCHING OF POLYMERS

Polymer materials have been used extensively for semiconductor devices and microsensors. Polymer films are used in microlithography for pattern definition. They are also used as micromolds for electroplating of metals or patterning of ceramic sensors. Dry etching allows the desired patterns to be controllably formed in polymers.[1-3] Multiple requirements are needed for polymer etching, including high etch rate, high selectivity, vertical profile, and large process latitude. Compared to other techniques of polymer patterning such as the synchrotron-based LIthographie, Galvanoformung, Abformung (LIGA) process[4] or conventional photolithography,[5,6] dry etching of polymers is a more flexible, controllable, and simpler technology.

3.2.1 Etch Characteristics for Thick Polymers

Thick polymer films are deposited and patterned by dry etching. For the etching of the thick Pyralin® PI-2611 polyimide, Ti and AZ5214 photoresist were used as the etch mask. The polyimide film was coated on 100 mm diameter wafers by multiple spinning. At a spin speed of 2 krpm for 30 s, the polyimide thickness was typically 8 μm with ±5% uniformity across the wafer. For 50 μm thick polyimide, 6 coatings were needed to build up the thickness. Soft bakes at 120°C for 3 min followed by 180°C for 3 min were applied in between each coating to partially drive out the solvent. After the last layer was spun, a final cure was carried out at 380°C for 1 hr. To avoid cracking of the thick polyimide film, the temperature of the hot plate was ramped up and down at a rate of 10°C/min to ensure slow solvent escape. A 300 nm thick Ti layer was evaporated on top of the polyimide as an etch mask for the subsequent O_2 plasma etching of polyimide. To define patterns on the Ti mask, 1.4 μm thick AZ5214 photoresist was spun on the Ti mask and patterned using conventional optical lithography. The Ti layer was etched by RIE in an Ar/BCl_3 plasma with the developed photoresist as an etch mask. In an O_2 plasma, Ti is oxidized to form TiO_x, which serves as a durable mask since TiO_x has a very low sputtering yield.

The effects of controllable parameters in the plasma system on polyimide etch rate, profiles, and selectivity were investigated to provide high aspect ratio features in the polyimide layer. Typical conditions used for polyimide etching were 750 W microwave power, 300 W rf power, 10 cm below the ECR source at 0.5 mTorr and 20 sccm O_2. The ECR source provides high dissociation efficiency due to the combination of magnetic confinement and microwave power, which results in a high etch rate. Low pressure reduces scattering of the reactive species and promotes anisotropic profiles and smooth surface morphology. Using the ECR source at low pressure allows high aspect ratio features to be generated with a vertical profile and high throughput.

Figure 3.2 shows the etch rate of polyimide and the self-induced dc bias ($|V_{dc}|$) as a function of microwave power. The samples were etched using 20 sccm O_2 at 10 cm below the ECR source. The chamber pressure was 1 mTorr at 100 W rf power and 0.5 mTorr at 300 W rf power. When the microwave power was increased from 250 to 1500 W, the etch rate of polyimide increased from 420 to 1035 nm/min at 100 W rf power, and from 509 to 1276 nm/min at 300 W rf power. The increase in etch rate is probably related to the enhanced dissociation of O_2 into reactive species when higher microwave power was applied. As microwave power was increased from 250 to 1500 W, $|V_{dc}|$ decreased from 159 to 113 V at 100 W rf power and from 320 to 264 V at 300 W rf power. From a power consideration, the rf power is roughly equal to the product of the sheath voltage and the ion current. Therefore $|V_{dc}|$, which is closely related to the sheath voltage, decreases as the ion density increases at constant rf power to the electrode. This is one of the major advantages of using the ECR source because high ion density can be generated with low $|V_{dc}|$. The low energy ions are beneficial in reducing surface damage and in enhancing selectivity. Etch profile was not influenced by microwave power as long as $|V_{dc}|$ was fixed. By keeping $|V_{dc}|$ at 200 V, the etch profile remained vertical even though the microwave power was varied from 250 to 1500 W. This result shows that ion energy is a more important factor for controlling etch profile than ion flux.

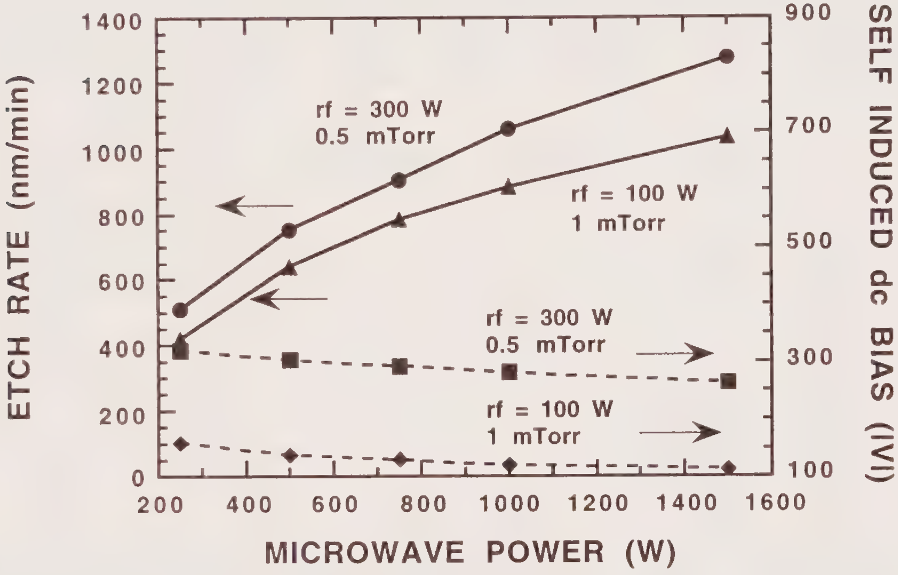


FIG. 3.2 Dependence of polyimide etch rate and self-induced dc bias on microwave power. The O_2 flow rate was 20 sccm at 10 cm below the ECR source. Pressure was 1 mTorr at 100 W rf power and 0.5 mTorr at 300 W rf power.

Comparisons of etch rate and selectivity between polyimide etching with an ECR source and RIE are shown in Figs. 3.3 and 3.4. For the polyimide etching with the ECR source, the microwave power was kept at 750 W, with a stage at 10 cm below the source, O_2 flow rate at 20 sccm, and chamber pressure at 0.5 mTorr. For RIE, the flow rate was 20 sccm at 3 mTorr pressure with the rf power varying from 50 to 300 W. This RIE condition was chosen to provide the lowest pressure possible in the plasma system so that the results could be compared with those obtained with the ECR source at low pressure. Figure 3.3 shows that the polyimide etch rate obtained using an ECR source is almost an order of magnitude faster than the etch rate obtained by RIE. This difference is mainly due to the higher ion densities generated by the ECR source.

Figure 3.4 shows the selectivity between polyimide and the Ti etch mask. Selectivity is much higher for samples etched in a plasma generated using an ECR source compared with RIE. With 50 W rf power, the selectivity was 3150 when etched in an ECR source and 528 when etched in the RIE system. The much higher selectivity obtained when the ECR source is used is related to the generation of high density ions at lower ion energy, whereas high ion density is obtained only by increasing rf power and hence the ion energy in the RIE system. For example, at 200 W rf power, $|V_{dc}|$ was 290 V for the plasma system with the ECR source and 500 V for the RIE system. The higher $|V_{dc}|$ caused a substantial increase in Ti etch rate and reduced the etch selectivity in the RIE system. It is also possible that the oxidation of Ti is more efficient in the ECR source, which results in oxide with lower sputtering yield. The faster etch rate and higher selectivity obtained in the O_2 plasma generated by the ECR source are particularly beneficial for etching high aspect ratio structures for microsensors.

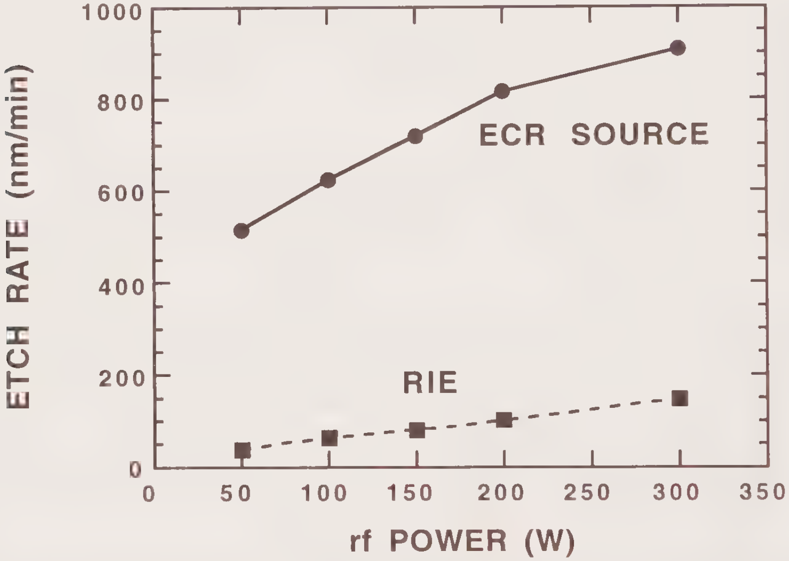


FIG. 3.3 Etch rate comparison between etching with an ECR source and RIE. For etching using the ECR source, microwave power was 750 W, rf power was 300 W, and O_2 flow rate was 20 sccm at 0.5 mTorr and 10 cm below the source. For RIE, the O_2 flow rate was 20 sccm and pressure was 3 mTorr.

The rf power not only controls the etch rate, it also has a strong influence on the etch profile. The etch profile can be related to the etch anisotropy. Anisotropy is defined as:

$$Anisotropy = 1 - \frac{Rate_{Horizontal}}{Rate_{Vertical}} \qquad (3.1)$$

With anisotropy = 1, there is no etching in the horizontal direction and the etch profile is vertical. Undercutting will reduce the anisotropy to be <1. The etch anisotropy was found to increase with rf power as shown in Fig. 3.5. At higher ion energy, the directionality of the ions has a weaker dependence on surface charge or field distribution across the wafer. The more directional ions impinging on the samples result in more vertical etch profile. The etch profile and surface morphology also depend on the pressure used during etching. More vertical etch profile and smoother surface morphology were obtained at lower pressure. This is related to the longer mean free path and reduced scattering at lower pressure. Since low pressure plasma down to 0.3 mTorr can be maintained with the ECR source, it is most suitable for generating a vertical etch profile for high aspect ratio resonant elements.

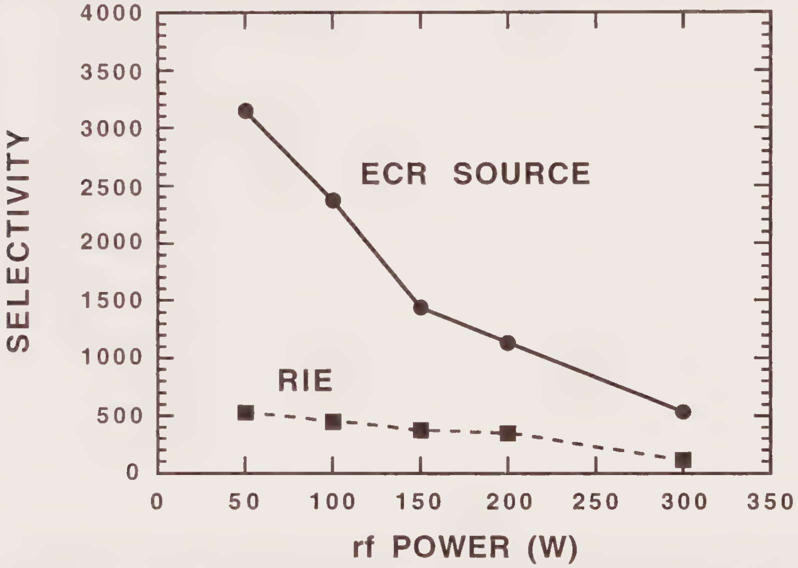


FIG. 3.4 Comparison of etch selectivity between polyimide and Ti. The etch conditions were the same as those used in FIG. 3.3.

3.2.2 Electroplating Ni through Polymers

To prepare the high aspect ratio polymer mold for electroplating, the dependence of etch rate on feature size was studied. In addition, high aspect ratio polyimide molds were formed with a thickness up to 32 μm and 1 μm gaps between microstructures. Figure 3.6 shows the decreased polyimide etch rate with reduced spacing between elements. The polyimide was etched down to a depth of 24 μm. The etch rate was normalized to the etch rate of 30 μm wide spacing. Both rf power and pressure can change the feature size dependent etch rate. At 0.5 mTorr and 300 W rf power, the normalized etch rate varies from 0.91 to 1.0 for spacing ranging from 4 to 30 μm. When the rf power was reduced to 50 W, the normalized etch rate decreased significantly, from 0.61 to 1.0 for the same range of spacing. The reduced etch rate for smaller features at lower rf power could be attributed to the less directional ions impinging on the samples. When the ion energy is lower, it is easier for the ions to be deflected by any perturbing electric field, resulting in lower ion flux at the bottom of high aspect ratio trenches. At higher pressure, the etch rate reduction for narrower spacing was more significant. The normalized etch rate for 4 μm spacing was 0.84 at 3 mTorr and 0.91 at 0.5 mTorr. The stronger dependence of etch rate on feature size is probably due

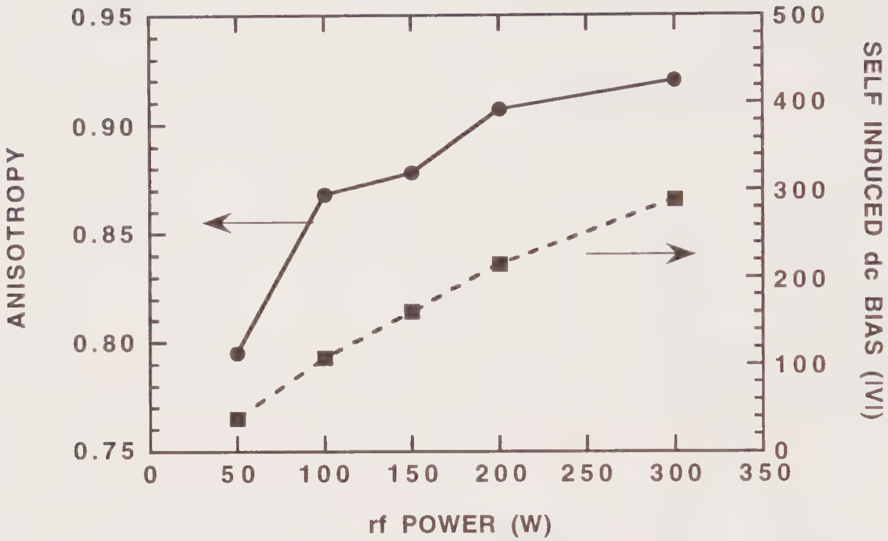


FIG. 3.5 Effects of rf power on polyimide etch anisotropy and self-induced dc bias. The oxygen flow was 20 sccm with 750 W of microwave power at 0.5 mTorr and 10 cm below the source.

to the additional scattering at higher pressure. The increased scattering makes it more difficult for the reactive radicals and the product species to enter or exit the high aspect ratio trenches.[7] Therefore, high rf power and low pressure are desirable to provide more directional reactive species to maintain a constant etch rate.

Figure 3.7 shows the high aspect ratio (32:1) resonant elements in polyimide that were etched with the ECR source. The polyimide microstructures are 32 μm deep with 1 μm spacing between elements. The surface is smooth, without any residue and the etch profile is vertical. A 25/5 nm thick Ti/Ni plating base was deposited below the polyimide mode by electron-beam evaporation. For Ni electroplating, the wafer was placed in an Ni plating solution (Barrett SN) and connected to the negative terminal of a current source. The positive terminal was connected to an Ni bar which was submerged in the solution and served to replenish the Ni ions in the solution that were deposited on the wafer. Typically, Ni was deposited at a rate of 400 nm/min using a plating current density of 10 mA/cm^2. After electroplating Ni through the polyimide mold, accelerometers with interdigitated resonant elements were formed as shown in Fig. 3.8. The gap between the Ni elements is 1.3 μm and the Ni is 14 μm thick. With the small spacing, high sensitivity is expected for the accelerometer.

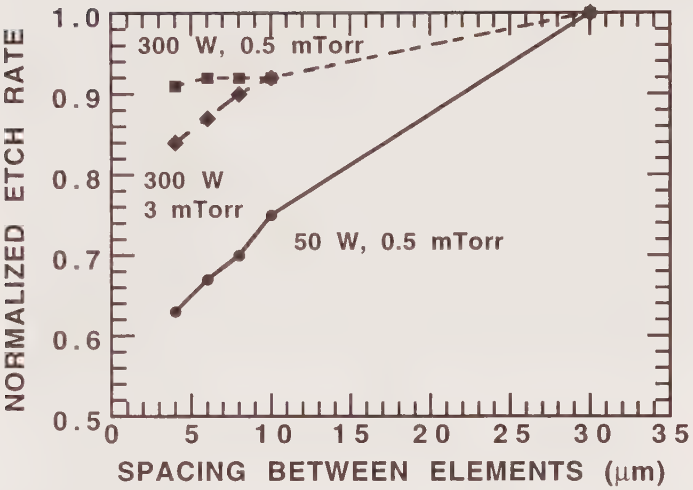


FIG. 3.6 Influence of feature size on polyimide etch rate. The O_2 flow rate was 20 sccm and the source distance was 10 cm with microwave power at 750 W and rf power at 50 and 300 W.

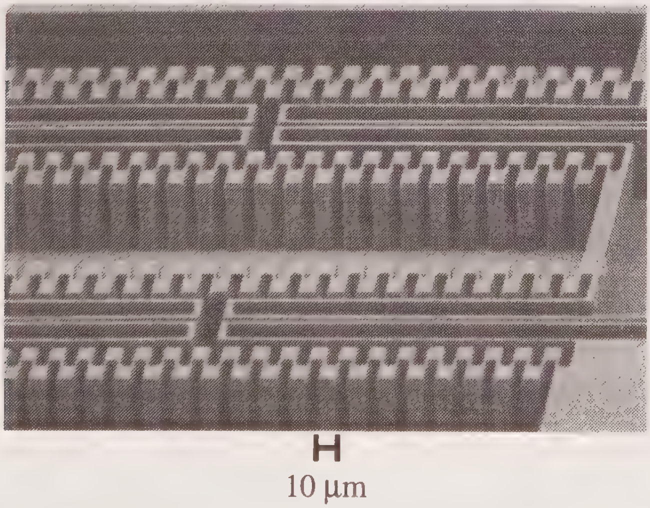


FIG. 3.7 High aspect ratio microstructures in polyimide. Etching was carried out with 750 W microwave power and -300 V at 0.3 mTorr, 12.5 sccm of O_2, and a 10 cm source distance. The polyimide etch depth was 32 µm with 1 µm spacing between resonant elements.

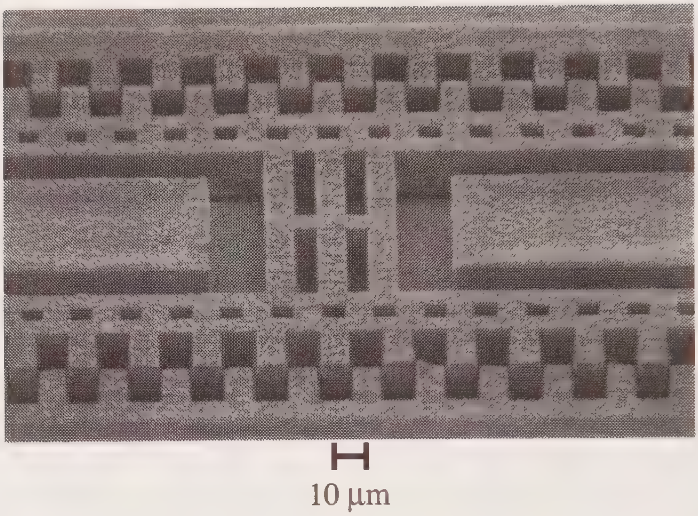


FIG. 3.8 Electroplated Ni resonant elements with aspect ratio of 11:1. The Ni was plated with a current density of 5 mA/cm^2 at 50°C.

Using the ECR source, polyimide was etched with a fast etch rate, high selectivity, vertical etch profile, and smooth surface morphology. The high density and low energy ions generated by the ECR source provide a polyimide etch rate of 0.91 μm/min and selectivity up to 3150. The low pressure plasma allows etching of polyimide with a vertical profile and smooth surface. Ti was found to have the highest etch selectivity over polyimide in the O_2 plasma. Microstructures that are 32 μm deep and 1 μm wide in polyimide have been fabricated. Electroplated Ni resonant elements that are 14 μm high with 1.3 μm spacing between elements have also been made.[8] Plasma etching with an ECR source has been demonstrated to have the capability of generating high aspect ratio resonant elements with small spacing for highly sensitive microsensors.

3.2.3 Micromolding of Ceramics

For high resolution ultrasonic imaging, high aspect ratio features in piezoelectric ceramics are needed.[9] Using dry etching technology to fabricate micromolds in polymer films,[10] it is possible to replicate deep features in the polymer micromold on conventional ceramic green tape made from submicrometer powder.

When an O_2 plasma is generated by an ECR source with high density and low ion energy, it provides a fast etch rate, residue-free surface, and an anisotropic etch profile for the fabrication of high aspect ratio polymer micromolds. Deep features in polyimide were etched to form the micromolds for piezoelectric ceramics. A 200 nm thick Ti layer was evaporated on top of the polyimide as an etch mask for the subsequent O_2 plasma etching of polyimide. Ti serves as a durable etch mask in an O_2 plasma since Ti is oxidized to form TiO_x, which has very low sputtering yield. Polyimide was etched using 750 W microwave power and 300 W rf power at 0.5 mTorr and a 20 sccm O_2 flow. The polyimide etch rate is 0.9 μm/min, with etch selectivity better than 1000:1 between polyimide and the Ti mask. Anisotropic profiles and smooth surface morphology can be obtained by adjusting the etch conditions. Figure 3.9 shows a 30 μm deep micromold in polyimide. The linewidth is 9 μm and the trench width varies from 4 to 25 μm. The etch profile is vertical and the etched surface is smooth.

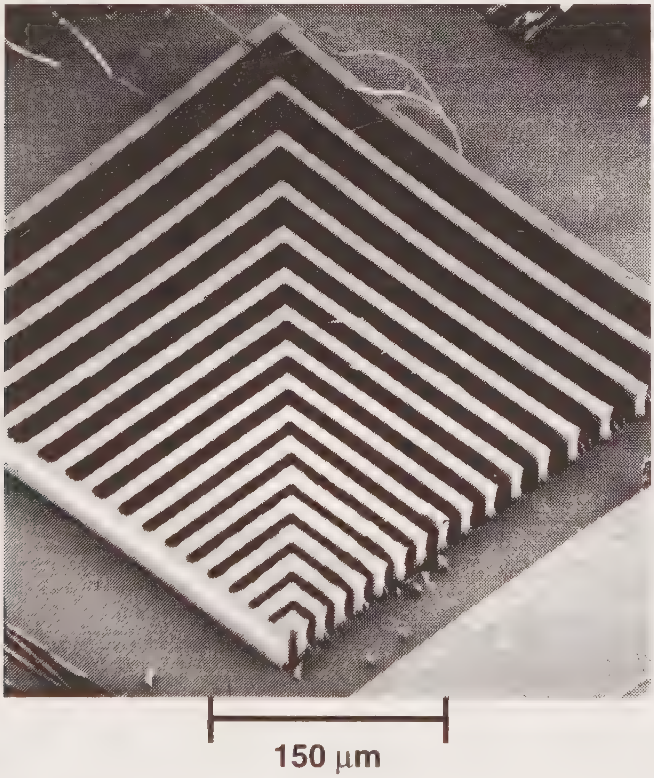


FIG. 3.9 A 30 μm deep polyimide micromold etched in an O_2 plasma. A 0.2 μm thick Ti was used as an etch mask with a selectivity better than 1000:1.

After the micromold is formed in polyimide, it is pressed onto a ceramic green tape made of submicrometer powder of a ceria-zirconia alloy. The ceramic was a Tosoh TZ-12CE, a 12 mol% CeO_2-88 mol% ZrO_2 powder with an average particle size of 0.3 μm. A hydraulic press with a pressure of 20 to 30 MPa and a dwell time of 30 s was used. After drying, the polyimide micromold was peeled from the tape, leaving an impression of the polyimide pattern on the ceramic tape. The patterned tape was slowly heated to 600°C to remove the binder and sintered at 1500°C for 60 min. The sintered ceria-zirconia has a grain size of 2 to 4 μm and a linear shrinkage of 20%. The ceramic replica of the polyimide mold is shown in Fig. 3.10. The linewidths in the ceramic vary from 7.5 to 22 μm and are the negative of the polyimide patterns. The fired ceramic patterns are 25 μm high, with an aspect ratio varying from 1 to 3. The spaces between the ceramic lines have a smooth bottom surface, reflecting the smooth surfaces of the polyimide patterns. Using the polyimide micromolds, fine features can be replicated in ceramic. This method should be suitable for micromolding any ceramic materials available as a sufficiently fine powder.

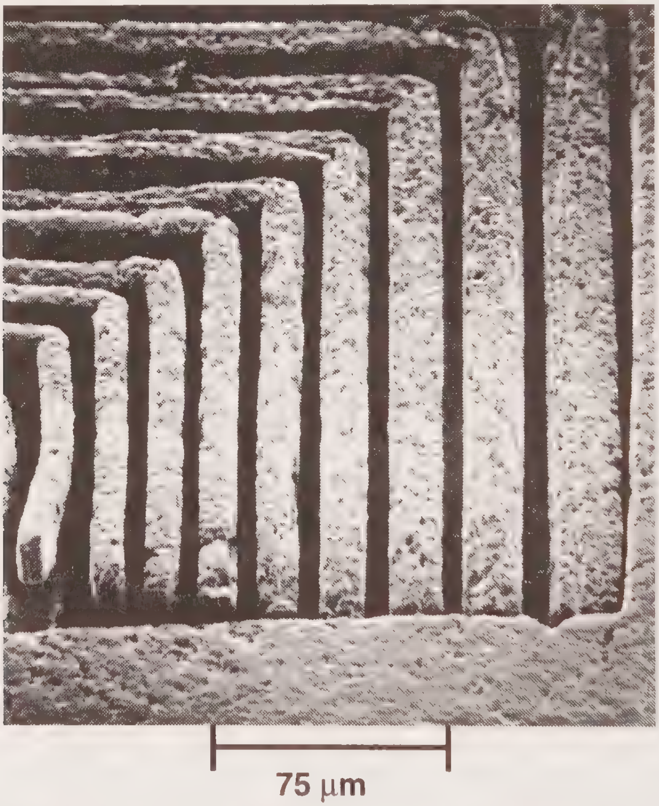


FIG. 3.10 Ceria-zirconia ceramic replica of the polyimide micromold. The patterns are 25 μm high after sintering at 1500°C for 60 min.

3.3 HIGH ASPECT RATIO SI MICROSTRUCTURES

Dry etching technology is essential for the fabrication of 3-dimensional microelectromechanical (MEM) devices in Si. Sensing elements with larger height and smaller gap are preferred because they enhance the devices performance significantly by increasing the electrostatic force between elements.[11] Reactive ion etching of Si for the fabrication of MEM devices has been reported previously.[12,13] Using RIE, problems related to slow etch rate, low selectivity, and complicated gas chemistry often limit the aspect ratio of microstructures to <10:1. Therefore, developing an etching technology that can provide high aspect ratio Si sensing elements is critical to MEM device fabrication since the devices performance can be improved substantially.

Using high density plasma sources, a high etch rate and a vertical profile for high aspect ratio microsensors can be obtained.[14,15] For etching Si sensing elements, Cl_2 instead of F-containing gases is used in the plasma system with an ECR source for anisotropic etching of Si. A vertical profile can be easily obtained in a Cl_2 plasma[16] and the process is more controllable, without the need for sidewall passivation using polymer forming gas chemistry or etching at low temperature. For deep Si etching, electroplated Ni was used as the etch mask. It has higher selectivity than SiO_2 and less intrinsic stress than evaporated Ni so that a thick layer can be deposited for the etching of high aspect ratio Si microstructures.

3.3.1 Etching of High Aspect Ratio Si Microstructures

In order to provide *in situ* monitoring for precise depth control during etching of Si, optical emission spectroscopy (OES) and mass spectrometry (MS) are used as the diagnostic tools. To relate the optical emission signals to the concentrations of reactive species, a small amount of Ar (5%) was added to the Cl_2 plasma as an actinometer to compensate for the variations of electron energy distribution in the plasma.[17] An Si emission at 288.1 nm normalized to a 294.2 nm Ar emission was used for monitoring since it provides a high signal to noise ratio. In addition, partial pressure from Si etch products in pure chlorine plasma was measured by MS. Although different Si etch products were found, $^{63}SiCl^+$ showed the highest intensity and good correlation with the Si etch rate. Figure 3.11 shows the etch rate, normalized Si emission intensity (Si intensity/Ar intensity), and $^{63}SiCl^+$ signal increase with microwave power. The plasma was excited at 100 W rf power at 3 mTorr with 20 sccm Cl_2 and 8 cm below the source. The etch rate increased from 84 to 406 nm/min when microwave power was increased from 50 to 1500 W. At the same time, the normalized Si emission intensity increased by 7 times and the $^{63}SiCl^+$ signal increased by 3 times. Signals from both OES and MS measurements show good correlation with the Si etch rate, indicating that a 288.1 nm Si emission and $^{63}SiCl^+$ signal are suitable for monitoring the etch process. The faster etch rate at higher microwave power can be attributed to higher flux of reactive species. The relatively slower increase of the $^{63}SiCl^+$ signal at high microwave power compared to Si emission intensity may be due to higher dissociation of the etch products. The etch profile was found to be more isotropic at higher microwave power, probably because of higher reactive species concentrations.

Besides using microwave power to control ion flux, rf power is used to control the ion energy through $|V_{dc}|$. As expected, the etch rate and selectivity are influenced by the rf power. For high aspect ratio etching of deep Si

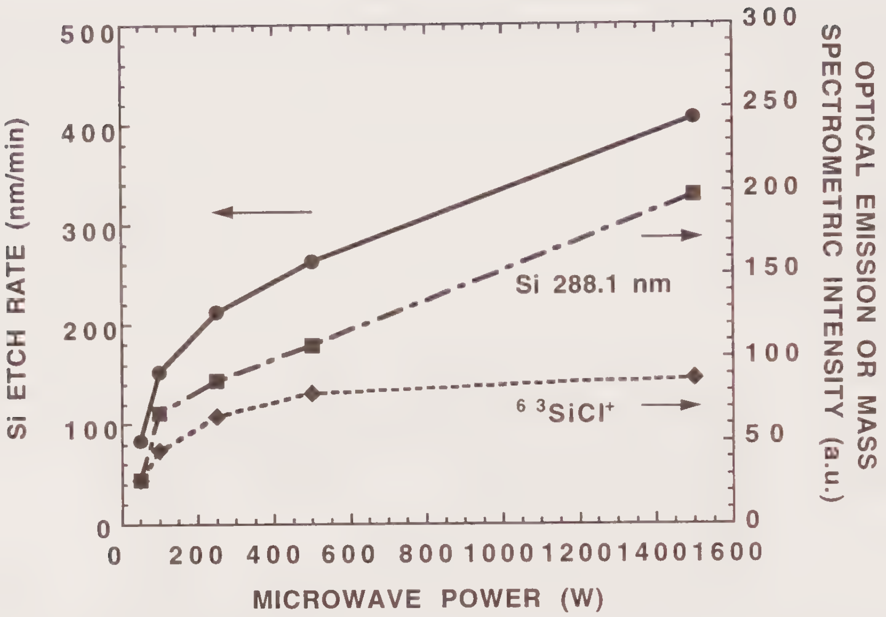


FIG. 3.11 Si etch rate, normalized Si emission signal at 288.1 nm (Si at 288.1 nm/Ar at 294.2 nm), and mass spectrometric $^{63}SiCl^+$ intensity as a function of microwave power. Samples were etched with 20 sccm of Cl_2 at 3 mTorr, 100 W rf power, and 8 cm below the ECR source.

microstructures, it is important that the Si etch rate be fast and the etch selectivity of Si over etch mask be high. The etch rate increased from 101 to 216 nm/min for rf power ranging from 50 to 300 W. Although high rf power is desirable for fast etch rate, the etch selectivity decreases with rf power. Figure 3.12 shows the durability of several etch masks in a Cl_2 plasma. Evaporated Ni was found to have higher selectivity than electroplated Ni, while thermal SiO_2 showed the lowest selectivity among the three etch masks. Evaporated Ni is not used as the etch mask because of its high intrinsic stress, which limits the thickness to <200 nm. The selectivity of evaporated and electroplated Ni decreased from 53 and 35 to 13 and 10, respectively, as rf power was increased from 50 to 300 W. Selectivity of SiO_2 was lower, decreasing from 9 to 7 within the same rf power range. This reveals that Ni is a more durable mask in Cl_2 plasma than SiO_2. There was a significant decrease in selectivity of Ni when rf power was increased from 50 to 75 W. This could be attributed to the threshold energy needed to sputter Ni. Although the selectivity is high at 50 W rf power, we did not apply this condition for our etches due to the slower etch rate. Within the range of rf power used, the Si etch profile is vertical and is independent of rf power.

Since feature size of different components on MEM devices varies substantially from 1 μm to several hundreds of micrometers, it is important to investigate how etch rate is influenced by feature size in order to provide

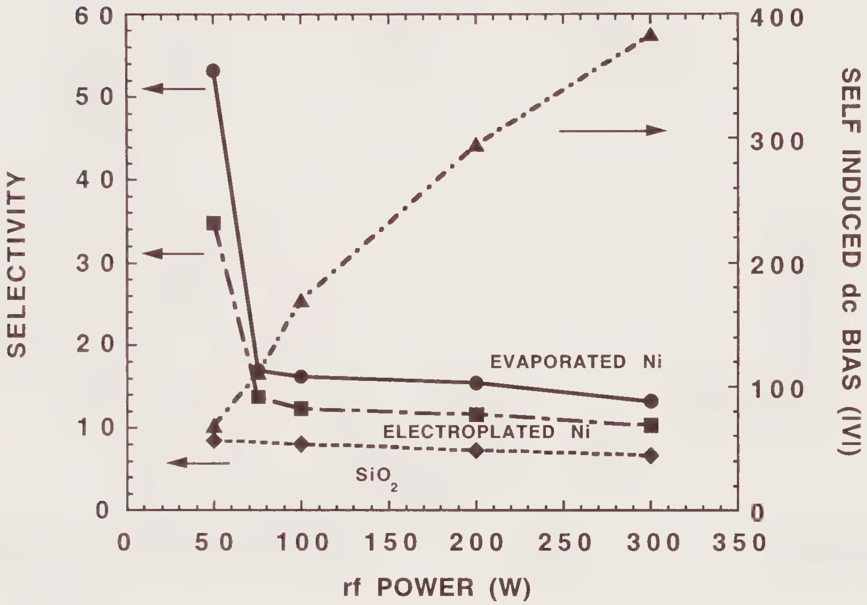


FIG. 3.12 Dependence of etch selectivity for evaporated Ni, electroplated Ni, and thermal SiO_2 etch masks on rf power. The Cl_2 plasma was generated by 100 W microwave power at 3 mTorr, a 20 sccm flow rate, and 8 cm below the ECR source.

sufficient overetch time. The dependence of etch rate on spacing between elements is shown in Fig. 3.13. The samples were 8 cm below the ECR source and were etched to a depth of 10 μm with 100 W microwave power, 100 W rf power, and 20 sccm of Cl_2 flow. The Si etch rate was normalized to the etch rate of 30 μm wide openings on the Si wafer. The normalized etch rate decreased from 1 to 0.95 and from 1 to 0.86 as trench width varied from 10 to 1.5 μm at 3 and 50 mTorr, respectively. The reduced etch rate for narrower features could be caused by decreasing concentration of reactive species at the bottom of trenches or by increasing difficulty of removing the etch products. At lower pressure, which corresponds to a shorter residence time at fixed flow rate, the etch rate reduction for high aspect ratio features is less significant. Reactive species encounter more collisions before reaching the samples at higher pressure. Therefore, the angular distribution of the reactive species is larger and more reactive species impinge on the sidewalls, resulting in a reduced etch rate for smaller features and a more prominent microloading effect. This is undesirable because a longer overetch time is needed for sensor fabrication. Since this effect could also be caused by the difficulty of removing etch products when aspect ratio is high, increasing flow rate to reduce the residence time could help to minimize the microloading effect. Nevertheless, we found there was only a slight reduction in microloading effect at higher flow rate under the etch conditions used.

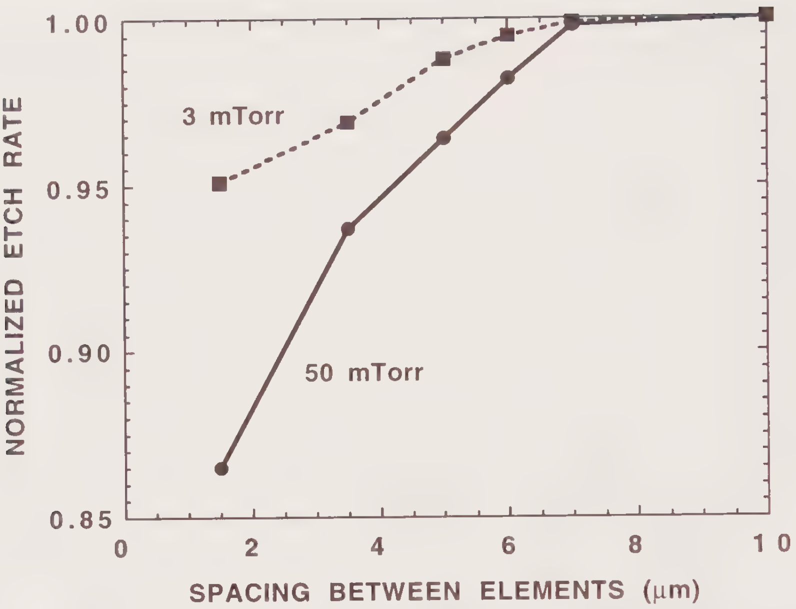


FIG. 3.13 Influence of spacing between elements on etch rate. Samples were etched with 100 W microwave power, 100 W rf power at 20 sccm Cl_2 and 8 cm below the source.

Si etch rate also depends on trench aspect ratio. Trench aspect ratio is defined as the ratio between the depth and the opening width of the trench. An average etch rate was obtained by dividing the trench etch depth by the total etch time, although the actual etch rate is faster at the initial stage and it slows down as the etch proceeds. For trenches with an aspect ratio >2, the average etch rate decreased linearly with aspect ratio, as shown in Fig. 3.14. The average etch rate was nearly constant for a trench aspect ratio ≤2. However, the etch rate started to decrease as the trench aspect ratio exceeded 2. The Si etch rate was 156 nm/min initially and it decreased to 119 nm/min for an aspect ratio of 34. For a trench aspect ratio ≥2, the Si etch rate can be fitted by R = 156 - A, where R is the average etch rate in nm/min and A is the trench aspect ratio.

Both etch rate and profile need to be controlled for Si microsensors with high aspect ratio elements. The effects of etch conditions on the etch rate and profile for deep Si microstructures are investigated. Pressure can have significant influence on the directionality of reactive species as well as the balance between etch and deposition reactions during dry etching. The Si etch rate for 2 µm wide openings was found to decrease from 151 to 124 nm/min as the pressure was increased from 3 to 30 mTorr, whereas the rate for 10 µm wide openings decreased from 163 to 141 nm/min. At lower pressure, the differences in the etch rate for the 2 and 10 µm openings are smaller. Lowering the pressure reduces the collisions of reactive species and etch products on the sidewalls. The ions are more directional and more ions can reach the bottom of narrow trenches since the number of collisions on the sidewalls due to diverged ions is reduced. It is also easier, at lower pressure, for the etch species and products to move through deep trenches.

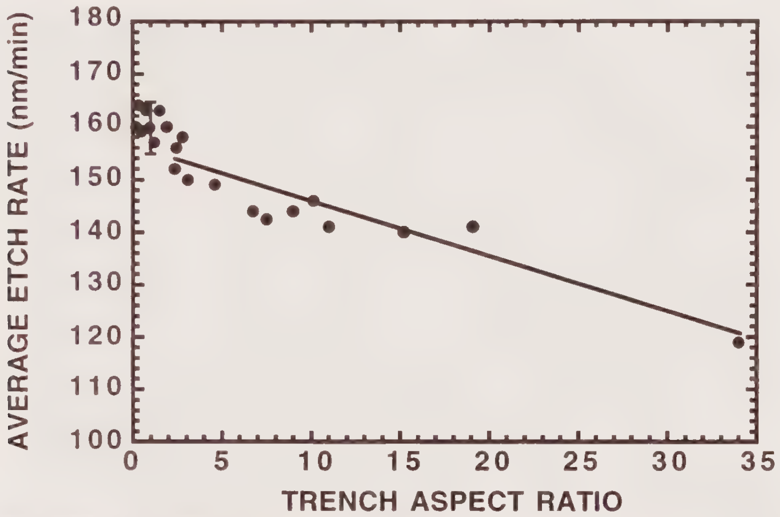


FIG. 3.14 Effect of trench aspect ratio on averaged Si etch rates for trenches with openings of 2, 3, 5, 20, and 50 µm. 100 W microwave power, 100 W rf power, 20 sccm of Cl_2, and an 8 cm source to sample distance at 3 mTorr were used.

Figure 3.15 shows the measured undercut widths as a function of pressure. The results are compared to those obtained using SPEEDIE simulation.[18] The samples were etched to a constant depth of 15 µm using 100 W microwave power, 100 W rf power, 20 sccm Cl_2, and an 8 cm source to sample distance. As the pressure was increased from 3 to 30 mTorr, the undercut width increased from 0.12 to 0.86 µm for 2 µm wide trenches and from 0.05 to 0.77 µm for 10 µm wide trenches. These results were similar to the undercut widths obtained using SPEEDIE simulations. For the Monte Carlo simulation, -160 V dc bias voltage, a 1 mm thick dark space sheath, and 1000 K of ion temperature were applied. Due to increased ion scattering at higher pressure, ions are less directional and more undercut is expected. The scattered species can impinge on the sidewalls and they can cause etching in the lateral direction. For narrower trench widths, collisions of diverged ions on the sidewalls are anticipated to increase further. However, with wider trenches, reactive species and etch products can move through the deep trenches more easily and with fewer collisions on the sidewalls by diverged particles. As a result, wider trenches have less undercutting and faster etch rates.

The effect of pressure on the directionality of ions was obtained from SPEEDIE simulations. For pressures ranging from 3 to 30 mTorr, the number of ions within 3° off normal of the wafer surface is approximately the same. Nevertheless, the number of ions at 30 mTorr is roughly an order of magnitude larger than that at 3 mTorr for larger incident angles. The wider ion angular distribution at higher pressure can cause increased etching in the lateral directions. The results from these simulations agree with the measured undercut widths.

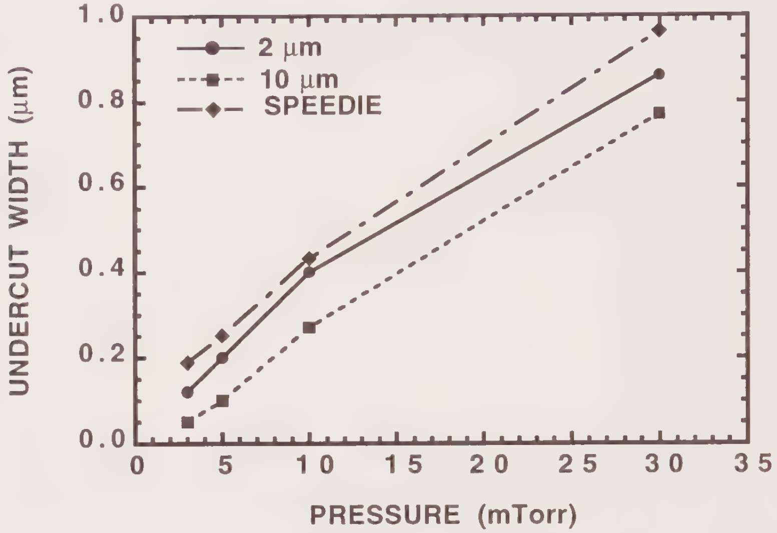


FIG. 3.15 Dependence of undercut width on pressure. The plasmas were generated with 20 sccm Cl_2, 100 W microwave power, 100 W rf power, and an 8 cm source distance. The etch depth was 15 µm.

By applying the optimized etch conditions, high aspect ratio microstructures in Si were obtained with 100 W microwave power, 100 W rf power, 20 sccm Cl_2 at 3 mTorr and 8 cm below the ECR source. Figure 3.16 shows Si resonators with an aspect ratio of 19. The etch depth is 28 µm, the spacing between elements is 1.5 µm, and the etched structures have a vertical profile and smooth morphology. The Si etch rate was 152 nm/min and it took 185 min to etch down to a depth of 28 µm. Even though etch rates can be increased by using higher microwave and/or rf power, they also result in an undercut etch profile and decreased selectivity, which are undesirable for etching of high aspect ratio microstructures. Further study is needed to increase the Si etch rate while still maintaining a vertical profile and high selectivity for etching sensing elements with a narrow gap and large depth. Despite the long etch time, these results show that by using a Cl_2 plasma generated by an ECR source, high aspect ratio etching of Si can be achieved and can be applied to the fabrication of MEM devices with higher sensitivity.

3.3.2 Deep Etch-Shallow Diffusion Technology

The bulk Si dissolved wafer process has been used to fabricate microstructures in single crystal Si.[19] In this process, a p^{++} etch stop layer is first formed by deep B diffusion. Microstructures are then fabricated by dry etching and released by wet etching of the Si substrate, stopping on the p^{++} layer. Since undercutting occurs during the conventional RIE step and a long diffusion time at high temperature is required to form a thick etch stop layer, the thickness of the microstructures fabricated by this technique is often limited to 10 μm.

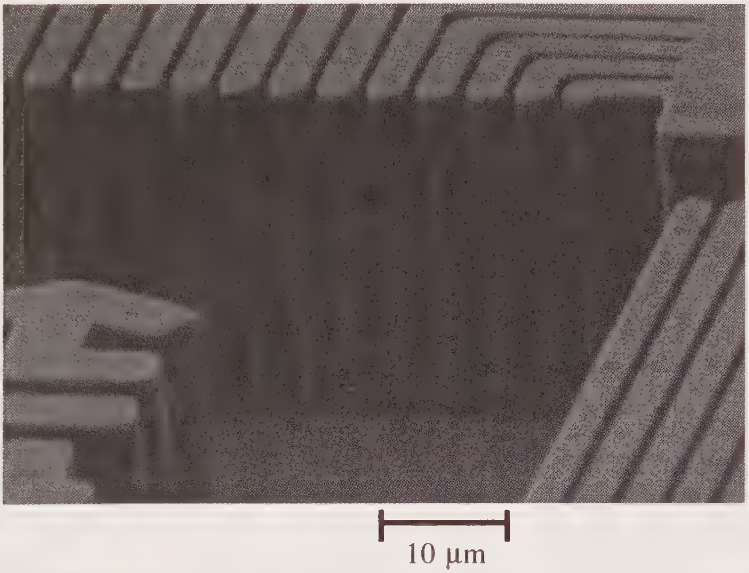


FIG. 3.16 Scanning electron micrograph of high aspect ratio Si microstructures etched down to a depth of 28 μm with 1.5 μm wide gaps. The etch conditions were 100 W microwave and rf power at 3 mTorr, with an 8 cm source distance and 20 sccm Cl_2 flow.

In this section, a novel deep etching and shallow diffusion process capable of fabricating much thicker Si microstructures with short diffusion time is described. This is accomplished by first dry etching thick Si microstructures with narrow gaps using an ECR source, followed by a shallow B diffusion to fully convert the etched microstructures to a p^{++} layer. A second dry etching step is then carried out to remove the shallow p^{++} layer at the bottom. Wafers are bonded to glass and wet etched in ethylenediamine pyrocatechol (EDP) and water solution to release the overhanging structures. This new etch-diffusion process allows high aspect ratio resonant elements to be fabricated in Si with a narrow gap and large thickness for high performance microsensors.

A schematic of the process sequence for this etch-diffusion process is shown in Fig. 3.17. The samples are made on B-doped (ρ=1-3 Ω-cm) Si wafers that are 10 cm in diameter and 550 μm thick. The wafers are first etched in a KOH solution with a 400 nm thermal oxide as an etch mask to form 4 μm deep bonding anchors. A plating base which consists of 25 nm of Ti and 5 nm of Ni is then evaporated on the recessed wafer. After patterning by photolithography, Ni is electroplated through the photoresist to form a dry etch mask for Si (Fig. 3.17a). The required thicknesses of photoresist and Ni are determined by the Si etch depth and etch selectivity between Si and Ni. In general, 2 μm thick electroplated Ni can hold up for 40 μm deep Si etching with the ECR source. The optimal etch condition for fabricating high aspect ratio microstructures is 100 W microwave power, 100 W rf power, an 8 cm source distance, 20 sccm Cl_2 at 3 mTorr and 25°C. Using this etch condition, resonant elements in Si with a vertical profile and smooth morphology are obtained for etch depths up to 40 μm (Fig. 3.17b).

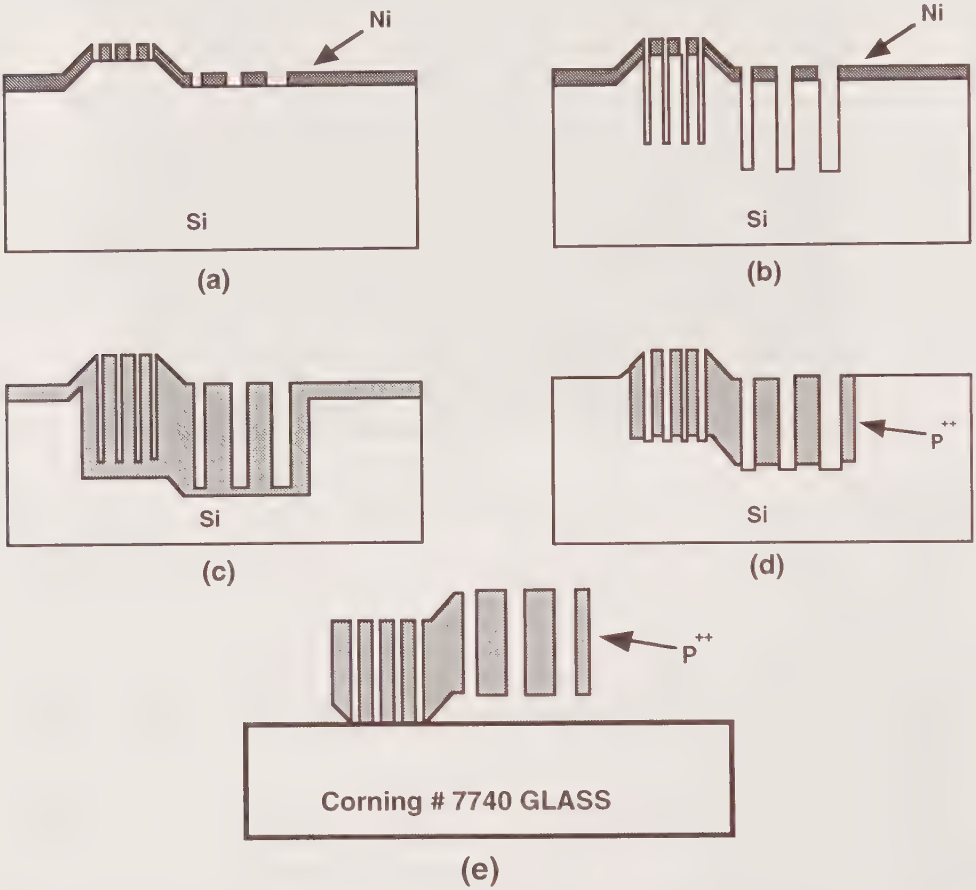


FIG. 3.17 Fabrication sequence for the deep etch and shallow diffusion process.

After the electroplated Ni etch mask is removed in HCl, a shallow B diffusion at 1175°C is then carried out to form a p^{++} layer in the resonators (Fig. 3.17c). The diffusion time is determined by the time needed to completely convert the entire width of the resonators from a lightly doped p-substrate to a p^{++} layer and not by the total depth of the resonators. In this case, very thick resonant elements can be fabricated by high aspect ratio Si etching and the thickness is not limited by B diffusion. Since B is diffused from both sides of a resonant element, only a short diffusion time is needed for B to diffuse to a distance equal to half of the element width. Due to the B diffusion around the bottom of the resonant elements, a second dry etching step is needed to remove this shallow p^{++} layer so that adjacent elements can be released and separated from one another. The same etch condition used in the first deep Si etch is applied to the second etch. A 50% overetch is typically used to ensure that the p^{++} layer at the bottom is completely removed (Fig. 3.17d). Finally the wafer is mounted upside-down on a carrier and thinned down to 300 μm by etching in HF/HNO_3 solution. The thinned wafer is anodically bonded onto Corning #7740 glass wafers. The bonding is carried out at 1000 V and 420°C for 10 min. The bonded Si microstructures are then released in EDP solution at 110°C for 5 hr to completely etch away the lightly doped Si substrate, leaving free standing, heavily B doped Si microstructures on a glass substrate (Fig. 3.17e).

Since precise lateral B diffusion from both sides of the resonant elements is needed to form the p^{++} layer, the diffusion time needed for lateral B diffusion was investigated. A patterned Si wafer was first etched down to 12 μm, followed by B diffusion at 1175°C. After the diffusion, the wafer was cleaved and etched in EDP solutions at 110°C for 2 hr. Since the lightly doped Si was removed in the wet etch process, the remaining heavily doped Si film could be used to determine the diffusion depth and profile. Figures 3.18(a) and (b) show the remaining p^{++} layers after the EDP etch. With 0.5 hr B diffusion, the diffusion depth was 3.1 μm, as shown in Fig. 3.18(a). This diffusion time was sufficient for the 4.5 μm wide element shown on the right, but a tunnel was formed for the 14 μm wide element on the left. A longer diffusion time of 6 hr allowed the 14 μm wide element to be completely converted to a p^{++} layer as shown in Fig. 3.18(b).

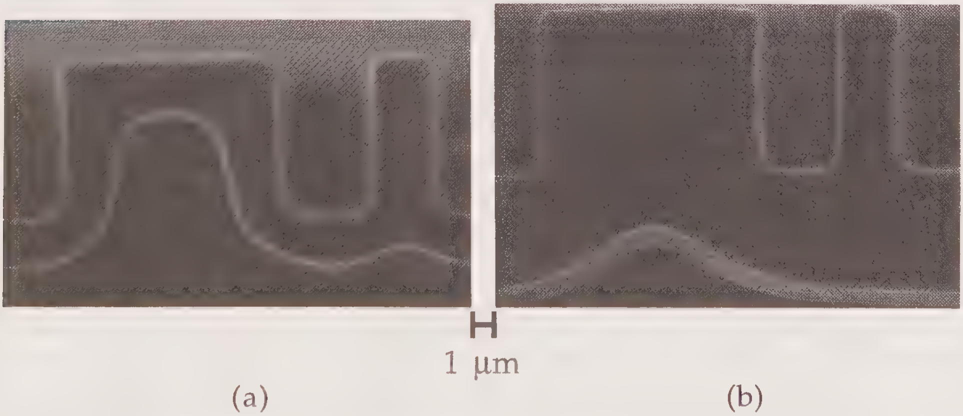


FIG. 3.18 Scanning electron micrographs of etch-diffused microstructures after B diffusion at 1175°C for (a) 0.5, (b) 6 hr.

The measured p^{++} film thickness for different diffusion times is shown in Fig. 3.19. This measured time dependent diffusion depth, T, can be fitted by the following:

$$T = 3.91t^{0.38} \tag{3.2}$$

where *t* is the diffusion time in hours. This expression can be used to estimate the B diffusion time needed for microsensors with different feature sizes. Since lateral diffusion occurs from both sides of the Si microstructures, the actual diffusion time needed is half of the time shown in Eq.(3.2).

Diffusion depth can also be calculated using the SUPREM3 simulation program. The simulated diffusion depths are shown in Fig. 3.19. An ambient B concentration of 10^{21} cm^{-3} and an etch stop at B concentration $\geq 2x10^{19}$ cm^{-3} were used for the simulations. The simulations show the general trends of diffusion depth as a function of diffusion time. However, this simulation program does not take into account the enhanced diffusion effect at higher doping concentrations. Therefore, a high ambient B concentration has to be used for the simulations to provide the best match to the measured diffusion depth, causing a discrepancy between the measured thicknesses and the simulated results. The SUPREM3 simulation shows that a 20-µm-thick p^{++} film in Si needs >50 hr of diffusion at 1175°C. Using this novel etch-diffusion process, the diffusion time can be reduced substantially, making it possible to fabricate high aspect ratio sensing elements in Si with large thicknesses.

3.3.3 Laterally Driven Comb Resonators in Si

In our mask design for resonators, the widest separation between openings is 12 µm. From the results shown in Fig. 3.19, 3.5 hr of B diffusion are needed to convert the widest elements to a p^{++} layer. Since a shallow p^{++} layer is formed at the bottom between the adjacent elements, a second etch step is needed to remove this layer. The second Si dry etching was carried out using the same etch condition as the initial high aspect ratio Si etch, but with no etch mask. Figure 3.20 shows the p^{++} Si microstructures that are 40 µm thick with a 2 µm wide gap between resonant elements. After the first Si etch in the Cl_2 plasma for 5 hr to get an etch depth of 40 µm, B was diffused at 1175°C for 5.5 hr to fully convert features that were 12 µm wide. A second Si etch for 1.5 hr with the same etch condition as the first was used to remove the shallow p^{++} layer at the bottom with a 50% overetch. Finally, a 30 min EDP wet etch was used to partially release the etch-diffused microstructures. With only 5.5 hr of B diffusion, 40 µm thick Si microstructures can be formed and released. This is a substantially shorter diffusion time compared to >50 hr needed if conventional diffusion and etch technology is used.

Figure 3.21 shows 30 µm deep resonant-based microsensors fabricated by this etch-diffusion process after being electrostatically bonded on glass. This is the first demonstration of bonded Si microstructures with a thickness as large as 30 µm. Since only a short diffusion time is used, it is expected that this p^{++} Si microstructure would have lower stress than those fabricated using conventional long B diffusion times. Buckled structures due to concentration gradient and nonuniform stress can be avoided.[20] Microsensors with thick sensing elements and small gaps in Si can be fabricated using this etch-diffusion process without being limited by the B diffusion time.

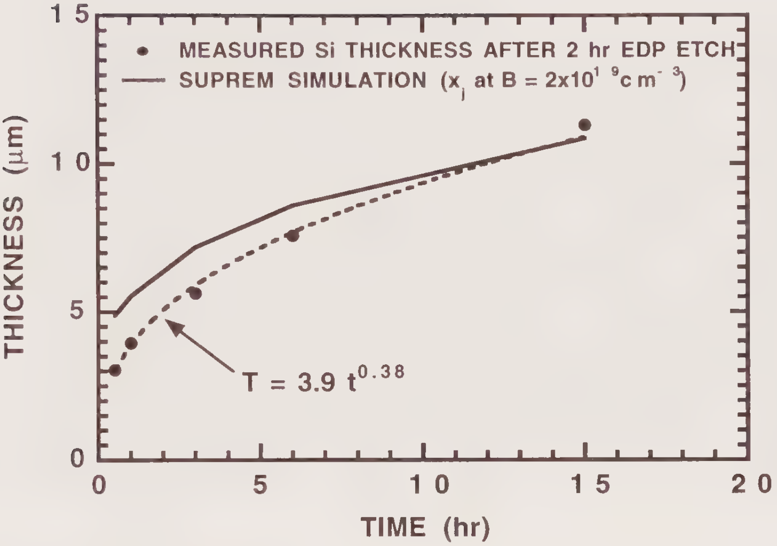


FIG. 3.19 Measured and SUPREM3-simulated p^{++} Si film thickness as a function of diffusion time.

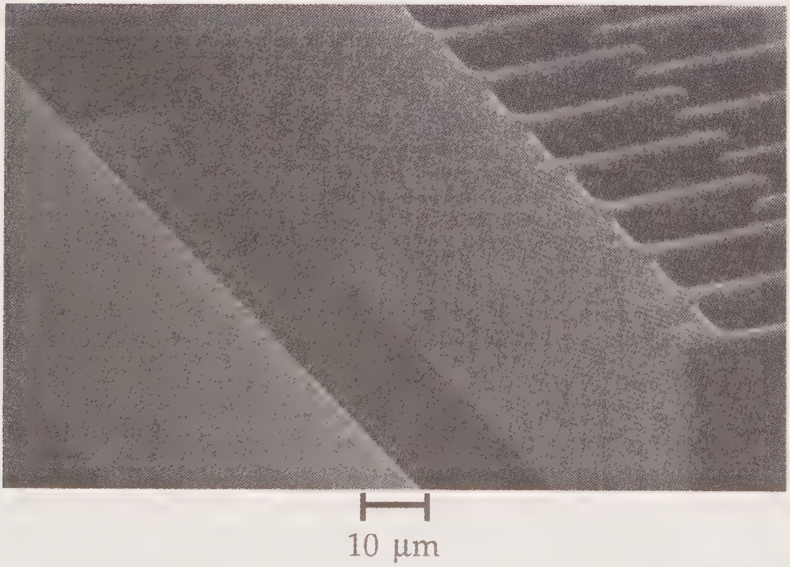


FIG. 3.20 Partially released overhanging Si microstructures that are 40 µm thick. The sample was first etched down to 40 µm, followed by 5.5 hr of B diffusion, a second Si etch to remove the 7.5 µm p^{++} layer at the bottom, and a 30 min EDP etch.

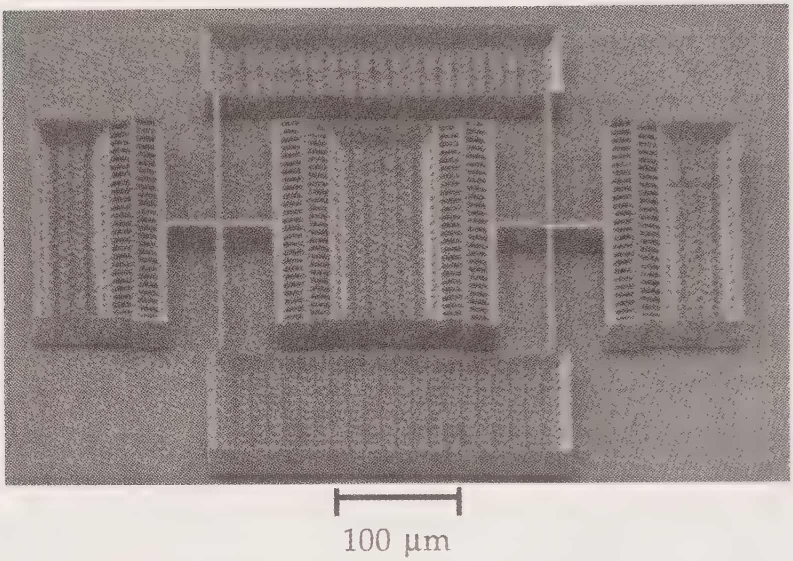


FIG. 3.21 30 µm thick Si resonant-based microsensors fabricated by the new etch-diffusion process after being bonded on a glass substrate and released in an EDP solution.

The electrical characteristics of a 35-µm-thick Si resonator fabricated using the deep etch and shallow diffusion process were measured. The resonant frequency was 2.49 KHz and Q was 23. The driving voltage required to achieve a 2.5 µm vibration amplitude was only 5 $V_{ac,p-p}$ superimposed on 5 V_{dc}. Thick resonant-based devices with a small gap can provide a large electrostatic force between resonating elements. In addition to a large force, thick devices also have the advantages of better mechanical strength and higher resistance to levitating force.[21]

3.4 SI FIELD EMITTERS WITH SHARP TIPS

Field emitter devices, in which electrons are emitted from a sharp tip under a strong electric field, have shown many promising applications, including flat panel displays, microwave tube cathodes, miniature microwave devices, ionization sources, and high power and frequency devices.[22-25] Field-emission flat panel displays are of interest because of their potential for ultrahigh resolution, small thickness, light weight, low power consumption, and good color quality. Furthermore, field emitter devices offer higher cutoff frequency, less temperature sensitivity, and better radiation hardness than other solid state microelectronic devices. Since their characteristics are mostly temperature-independent, field emitters are excellent structures for microsensor applications. They can also be used as electron sources for scanning electron microscopes, scanning tunneling microscopes, and electron beam lithographic systems. Since electron emission strongly depends on the electric field, and the electric field scales with emitter sharpness, field emitters should be as sharp as possible. Sharp emitters also make it possible for electron emission devices to operate at low voltage. Emitters with larger height have lower intrinsic capacitance and higher cutoff frequency. Therefore it is desirable to have emitters with sharp tips, large height, and high packing density.

There are several techniques for fabricating field emitter tips. The two most frequently used techniques include evaporation through small openings[26] and etching of Si emitters.[27] The first approach usually requires tilting and rotating the sample during evaporation. Wet chemical etching is typically used for the second approach. Emitter tips are formed along <111> Si planes because etching is dependent on crystal orientation. Dry etching of Si provides more flexibility for the fabrication of field emitter tips. Compared to wet etching of Si tips, dry etching provides more control of the etch profile.[28,29] Sharper and taller tips with higher packing density can be fabricated by dry etching.

3.4.1 Control of Emitter Profile by Dry Etching

Si emitter tips were etched in a Cl_2 plasma generated by an ECR source. SiO_2 was used as the etch mask and a mask erosion technique was employed to control the Si emitter profile. The SiO_2 mask was eroded during the etch and the Si emitter profile was determined by controlling the lateral etch rate of the SiO_2 mask and the vertical etch rate of Si. By using SiO_2 masks with different sidewall angles and adjusting the etch conditions, the lateral etch rate of the SiO_2 mask was changed and Si emitter tips with different sidewall angles were fabricated.

It was found that the lateral etch rate of SiO_2 greatly depends on the initial mask profile. Lateral and vertical etch rates of the SiO_2 mask, as well as Si profile angle (q) could be adjusted by the mask profile and the etch conditions. The dependence of a lateral SiO_2 etch rate on the initial mask profile is shown in Fig. 3.22. With 250 W microwave power and 200 W rf power applied to a Cl_2 plasma at 2 mTorr and 20 sccm Cl_2, Si samples were etched to a depth of 1.8 µm. The vertical etch rates of Si and SiO_2 were 250 and 63 nm/min, respectively, at the above conditions and did not change with the mask profile. As shown in Fig. 3.22, the lateral etch rate of SiO_2 decreased from 363 to 150 nm/min as the slope of the SiO_2 mask increased from 17 to 45°. The results are consistent with the angle dependent sputtering yield,[21,22] showing an increase in lateral etch rate at lower SiO_2 mask angles. However, as the mask slope increased from 35 to 45°, the lateral etch rate decreased only from 156 to 150 nm/min. Due to the decrease in the lateral SiO_2 etch rate at a larger sidewall angle (ϕ), the Si profile is less tapered as the initial mask profile becomes more vertical. The slope of the Si profile increased from 52 to 78° as the initial mask angle increased from 17 to 45°. Although the lateral etch rate of SiO_2 decreases more slowly at $\phi > 35°$, the Si profile (θ) still increases considerably. This is expected, since at more vertical

Si profiles (higher θ), small changes in the lateral etch rate will result in considerable variations in the Si profile slope.

In order to fabricate Si emitter tips using this mask erosion technique, the SiO_2 mask should hold up during the etch and be completely eroded at the end of the etch. The mask should be thick enough so that it is not etched away vertically before it is eroded laterally. Therefore the width and the thickness of the mask should be chosen in accordance with the vertical and lateral etch rates, as well as the final basewidth of the emitter tips.

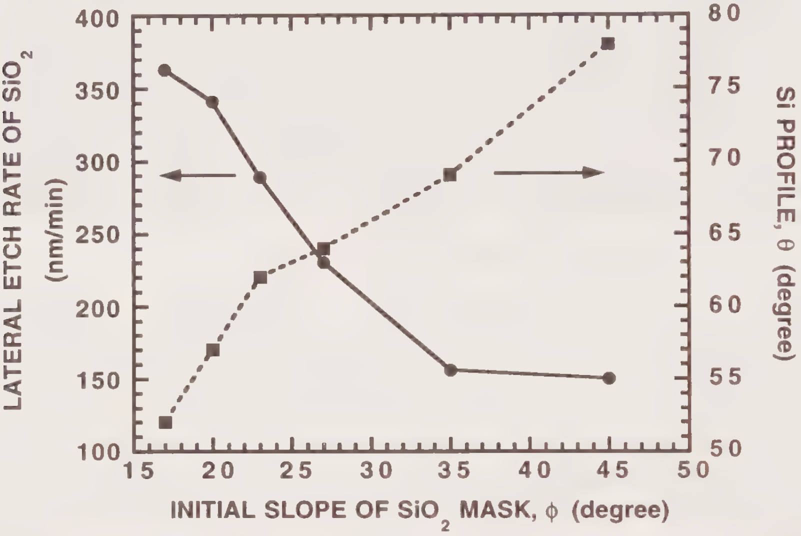


FIG. 3.22 Dependence of SiO_2 lateral etch rate and Si etch profile on the SiO_2 mask slope. The etch conditions were 250 W microwave power, 200 W rf power, and 20 sccm of Cl_2 at 2 mTorr.

The rf and microwave power are used to control the energy and density of ions in the plasma. The energy of ions in the plasma is adjusted by varying the rf power applied to the stage, and is related to $|V_{dc}|$ across the plasma sheath. As rf power was increased from 50 to 400 W, $|V_{dc}|$ increased from 97 to 368 V and the lateral etch rate of the SiO_2 mask with an initial sidewall angle (ϕ) of 45° increased from 54 to 200 nm/min, as shown in Fig. 3.23. The vertical etch rate of Si also increased from 170 to 266 nm/min. The increases in the etch rates are related to the energy of ions in the plasma, which increases with rf power. Although the ion energy and the etch rates increased with rf power, the Si etch profile did not change significantly. It varied slightly from 80 to 74°, showing only a 6° change with an eight fold increase in rf power. Similar results were obtained when SiO_2 masks with $\phi = 17°$ were used. The Si profile angle only decreased from 55 to 47° as rf power was increased from 50 to 500 W. The slope of the Si profile is determined by the SiO_2 lateral etch rate and the Si vertical etch rate. Since both etch rates show a similar increase with rf power, no significant change in Si profile is observed. Hence, rf power can be used to achieve the desired etch rate without changing the Si profile. However, energetic ions generated at higher rf power could introduce surface damage and degrade device performance. Therefore, very high rf powers should be avoided.

The Si emitter profile also remained nearly unchanged when microwave power was varied. As microwave power was increased from 50 to 800 W, the slope of the Si profile only decreased from 80 to 75° for an SiO_2 mask with $\phi = 45°$. Both the Si and SiO_2 etch rates increased with microwave power due to higher ion density in the plasma. Similar to the results for rf power, microwave power can be used to change the ion density and adjust the etch rate without changing the Si etch profile.

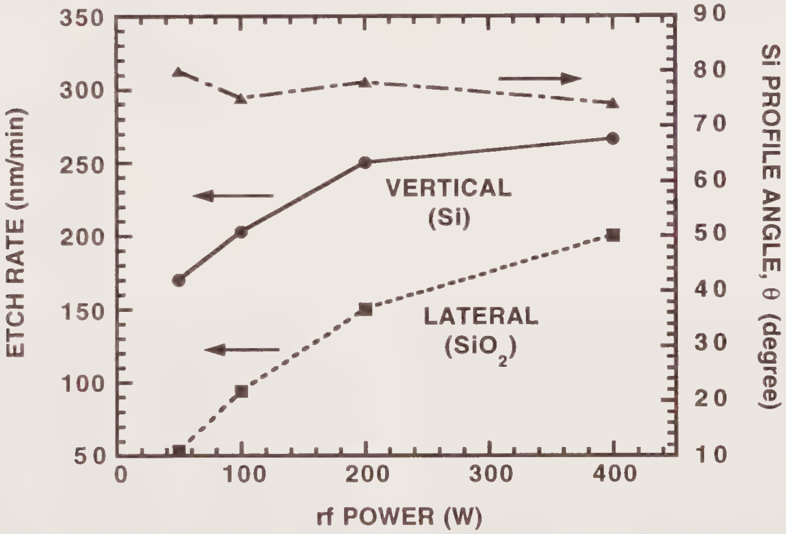


FIG. 3.23 Etch rate and Si profile variations with rf power. Samples were etched in a Cl_2 plasma at 2 mTorr with 250 W microwave power.

The effects of pressure on Si etch rate are shown in Fig. 3.24. The pressure was increased from 0.5 to 10 mTorr in a plasma with 250 W microwave power, 200 W rf power and 20 sccm Cl_2. The lateral etch rate of SiO_2 decreased from 414 to 144 nm/min and from 173 to 63 nm/min for the SiO_2 masks with $\phi = 17°$ and 45°, respectively. The vertical etch rate of Si only changed from 169 to 206 nm/min in the same range. These different variations in etch rates could be attributed to the scattering of the ions, which increases with pressure. As pressure increases, energetic ions will lose their directionality and energy due to increased scattering. These cause the lateral etch rate to decrease with pressure since the sputtering yield depends on angle and ion energy. The effects are more substantial for features with smaller ϕ due to angular-dependent sputtering yield. Therefore, increasing pressure will reduce the directionality of ions and change the effective incident angle between ions and etch surface, resulting in lower sputtering yield and lower lateral etch rate. These effects can be seen in Fig. 3.24. The lateral SiO_2 etch rate decreased as pressure was increased, and this reduction was much more substantial for SiO_2 masks with $\phi = 17°$ compared to $\phi = 45°$. The vertical Si etch rate first increased, then decreased slightly at high pressure. Since the vertical etch rate does not depend strongly on the directionality of ions, the decrease in the vertical etch rate with pressure is much less than the reduction in lateral etch rate. Because the variations in the SiO_2 mask lateral etch rate with pressure are substantially different from the variations in the vertical Si etch rate, a large range of Si etch profiles is obtained with changing pressure. As Fig. 3.25 shows, for the masks with $\phi = 17°$, the slope of the Si profile increased from 42 to 68° as the pressure was increased from 0.5 to 10 mTorr. The slope of Si profiles with $\phi = 45°$ increased from 72 to 85° when pressure was increased from 0.7 to 10 mTorr, indicating that more vertical profiles can be obtained at higher pressures.

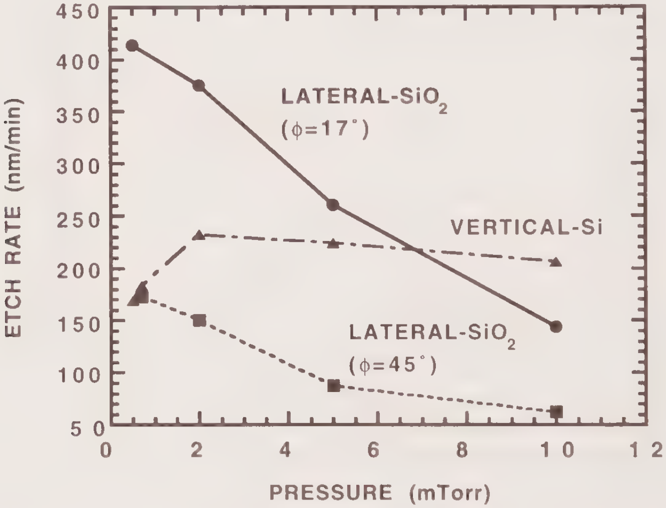


FIG. 3.24 Variation of Si and SiO_2 etch rates with pressure. The etch conditions were 250 W microwave power, 200 W rf power, and 20 sccm of Cl_2.

Optimized etch conditions were used to fabricate field emitter arrays in Si. Microwave and rf powers were adjusted to get the desired etch rate and basewidth, while the SiO_2 mask profile and pressure were used to control the slope of Si emitters. Figure 3.26 shows Si emitter tips etched with 250 W microwave power, 400 W rf power, and 20 sccm Cl_2. For the emitter shown in Fig. 3.26(a), an SiO_2 mask with $\phi = 23°$ was used and the pressure was 0.9 mTorr during the etch. Since the lateral etch rate of SiO_2 is high for low ϕ and at low pressures, more tapered Si profiles (smaller θ) are developed. The slope of emitters, θ, is 55° in this case. The emitter shown in Fig. 3.26(b) was etched at 10 mTorr using an SiO_2 mask with $\phi = 38°$. This resulted in a much more vertical Si profile with a sidewall angle of $\theta = 82°$. This is due to the lower lateral etch rate of SiO_2 at higher ϕ and higher pressures.

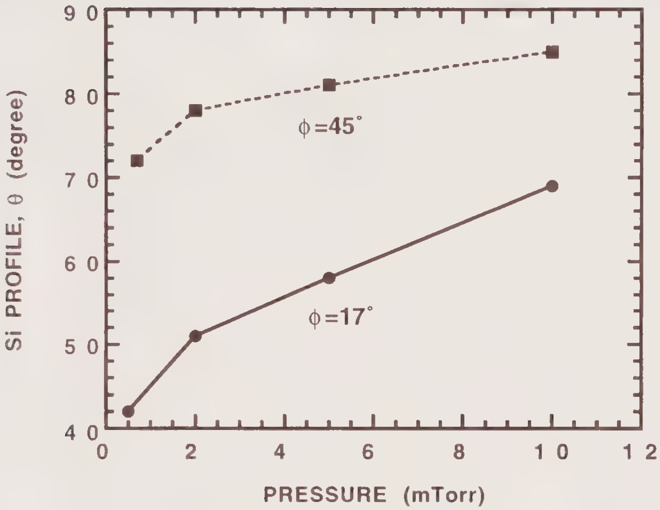


FIG. 3.25 Dependence of Si etch profile on pressure. The samples were etched under the same conditions as shown in FIG. 3.24.

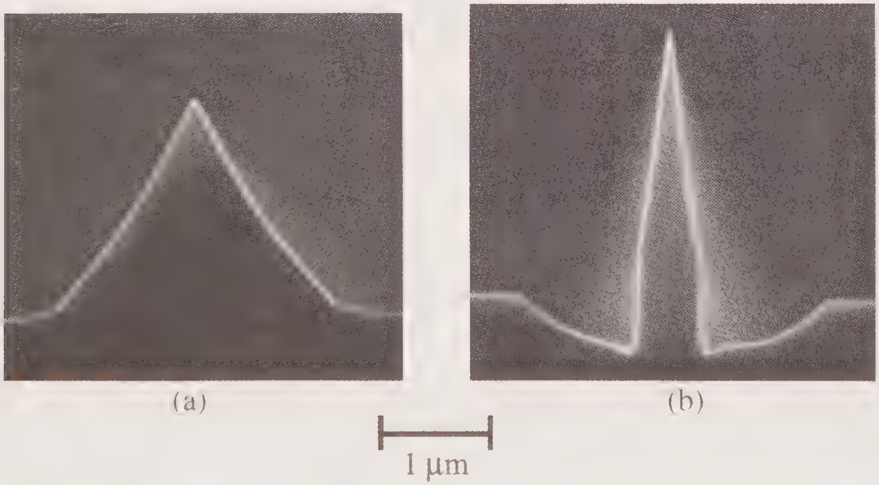


FIG. 3.26 Scanning electron micrographs of Si emitters. The etch conditions were 250 W microwave power, 400 W rf power, and 20 sccm Cl_2. The pressure was 0.9 mTorr and the SiO_2 mask profile was 23° for (a). For (b), the pressure was 10 mTorr and the SiO_2 mask profile was 38°.

3.4.2 Sharpening Field Emitter Tips

Si emitter tips were formed in a Cl_2 plasma generated at 5 mTorr with 20 sccm Cl_2, 250 W microwave power, and 200 W rf power. A 1-μm-thick thermal oxide patterned in a CF_4 plasma was used as the etch mask and the mask was eroded to generate Si emitter tips. The emitter tip radii were measured by scanning electron microscopy to be 67±3 nm right after the SiO_2 mask was completely eroded.

Further dry etching in the Cl_2 plasma was used to sharpen the Si emitter tips. Figure 3.27 shows the effects of etch time and pressure on the emitter tip sharpness. The emitter tip radius, defined as the radius of curvature at the tip apex, was measured before and after sharpening using scanning electron microscopy. The variation in the tip radius was found to be less than ±3 nm in all measurements, as indicated by the error bars in Fig. 3.27. The tip radius was 67 nm before sharpening began and it was reduced to 30 nm after 2 min of etching at 0.7 mTorr. Further etching for 5 more min only reduced the tip radius to 22 nm. The sharpening effect at 0.7 mTorr could be related to the higher sputtering yield at sloped surfaces.[30] As a result, the emitters develop more vertical profiles close to the tip apex. As the etch progresses, the emitters become sharper near the top. Since the sputtering yield decreases for more vertical profiles, even though the emitter tips become sharper with longer etch time, the amount of sharpening is less at longer times.

Sharpening the emitter tips by dry etching is less efficient at higher pressure. As shown in Fig. 3.27, the tip radius decreased from 67 to 62 nm after 7 min of etching at 5 mTorr. As the pressure in the plasma increases, the directionality and energy of ions impinging on the etch surface will decrease due to more collisions in the plasma. The ions will have a broader angular distribution and will strike the etch surfaces at a wider angular range at higher pressure. As a result, the tip apex will be etched more isotropically and the sharpening effect will be decreased. At 5 mTorr, the emitter sharpness did not change significantly with longer etch time. It only reduced the height of the emitters.

Figures 3.28(a) and (b) show scanning electron micrographs of emitter tips before and after being sharpened by dry etching in the Cl_2 plasma. The emitter tip radius was 67 nm before sharpening as shown in Fig. 3.28(a). After the emitters were etched for 7 min at 0.7 mTorr, the tip radius decreased to 22 nm, as shown in Fig. 3.28(b). The sloped area close to the emitter apex etched faster due to angular-dependent sputtering yield and a more vertical profile near the tip apex developed. Therefore, sharper emitter tips with smaller tip radius can be formed by controlling the etch time after the oxide mask is completely eroded.

Optical emission spectroscopy and mass spectrometry were used to monitor the fabrication and sharpening of Si field emitter tips. Samples were etched in a Cl_2 plasma at 0.7 mTorr with 250 W microwave power, 200 W rf power, and 20 sccm of Cl_2. The optical emission signal of Si at 288.1 nm and the mass spectrometric signal for $^{63}SiCl^+$ were monitored during the etching of Si emitters with SiO_2 as the etch mask. The results are shown in Fig. 3.29. Before the SiO_2 mask was completely etched away, the Si signal increased

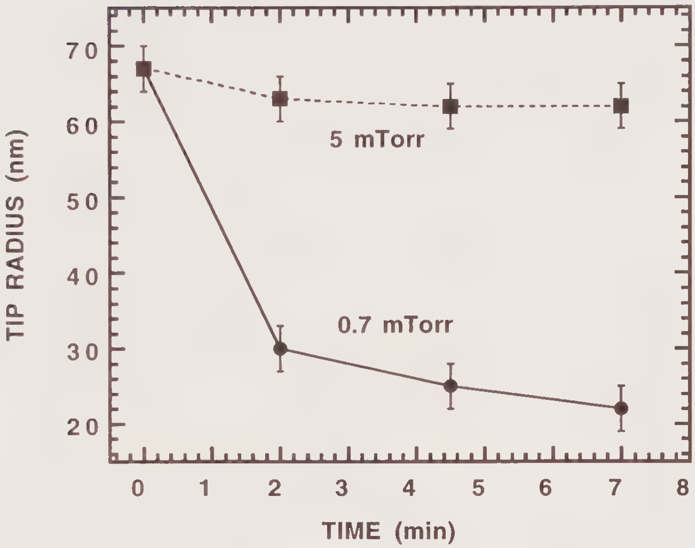


FIG. 3.27 Sharpening Si field emitter tips by dry etching. Samples were etched in a Cl_2 plasma with 250 W microwave power, 200 W rf power, and 20 sccm of Cl_2. The source to stage distance was 10 cm. Error bars show that the variations in the tip radius were less than 3 nm.

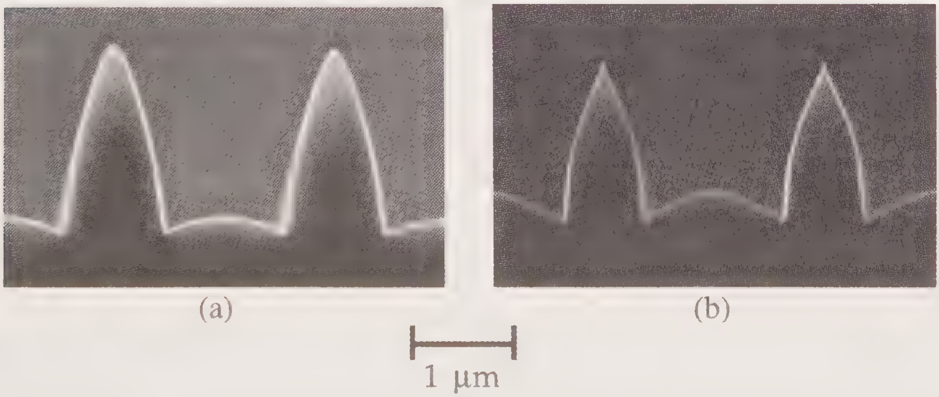


FIG. 3.28 Scanning electron micrographs of Si emitter tips sharpened by dry etching in a Cl_2 plasma with 250 W microwave and 200 W rf power. (a) Emitter tips before sharpening with tip radius of 67 nm and (b) emitter tips after sharpening at 0.7 mTorr for 7 min. The tip radius is reduced to 22 nm.

slightly due to gradual erosion of the SiO_2 mask as etching progressed. After 14 min of etching, the oxide mask was completely removed and the Si emitters had a tip radius of 67 nm. Once the oxide mask was eroded completely and Si tips were formed, sharp increases in the Si signals were observed since the entire wafer consisted of Si at this time. The increases in Si signals indicate the complete removal of the SiO_2 mask, after which the tip sharpening process starts. The tips were sharpened as the samples were etched further at 0.7 mTorr. The tip radius decreased from 67 to 22 nm after an additional etching of 7 min. The optical emission and mass spectrometric signals can be used for *in situ* monitoring when the emitter tips are sharpened. The etching can be stopped once the desired tip radius is achieved.

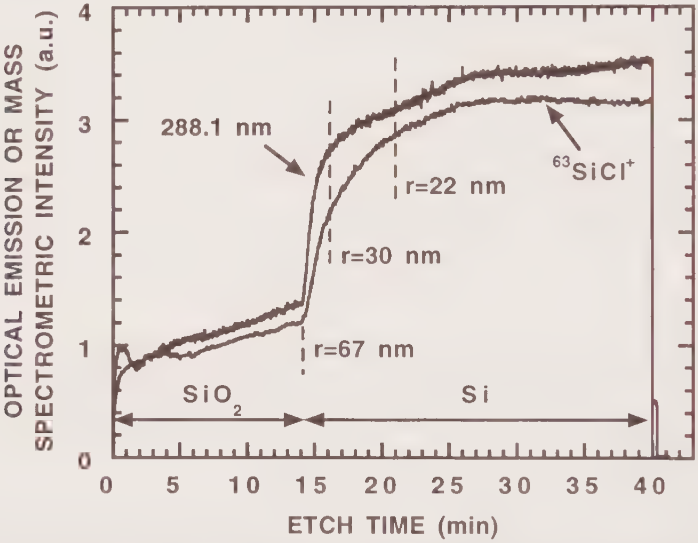


FIG. 3.29 Normalized Si emission intensity at 288.1 nm and mass spectrometric signal for $^{63}SiCl^+$ as a function of etch time. The etch conditions were 250 W microwave power, 200 W rf power, 0.7 mTorr, a 10 cm source distance, and 20 sccm of Cl_2.

Low temperature oxidation in an O_2 plasma was also used to sharpen Si emitter tips. Using the ECR source, a high density O_2 plasma was generated at 0.3 mTorr with 10 sccm of O_2 and 1000 W microwave power applied to the source. No rf power was applied and the stage temperature was kept at 30°C. The distance between the ECR source and the stage was set at 8 cm. The efficient O_2 dissociation at low pressure in the ECR source allows a high density plasma to be created for a high oxidation rate at a low temperature.[31] No rf power was used to avoid sputtering of the Si emitters by energetic ion bombardment. After oxidation in the O_2 plasma for 70, 200, and 500 min, oxide thicknesses of 51, 98, and 257 nm were grown, respectively. The tip radius was measured by scanning electron microscopy before oxidation and after the plasma oxide was removed in buffered HF. The reduction of tip radius after plasma oxidation and oxide removal is shown in Fig. 3.30. The error bars in Fig. 3.30 show that the variation in the tip radii was less than 3 nm. Before oxidation, Si emitters with tip radii of 67±3 nm were fabricated by dry etching in the Cl_2 plasma at 5 mTorr. The tip radius decreased to 25 nm after 51 nm of plasma oxide was grown and subsequently removed. This corresponds to ~42 nm reduction in the tip radius. Growing a thicker oxide up to 257 nm reduced the tip radius to 8 nm. Figure 3.31 shows the scanning electron micrographs of emitter tips sharpened by plasma oxidation. For a plasma oxidation time of 70, 200, and 500 min, the Si field emitter tip was sharpened to 25, 18, and 8 nm, respectively. The results suggest that Si field emitters can be sharpened efficiently by plasma oxidation.

Thermal oxidation was also used to sharpen Si emitter tips. To compare the results with plasma oxidation, dry oxides with similar thicknesses were grown at 950 and 1100°C. Dry oxidation was chosen to provide a better control over the oxidation rate for thin oxides. As shown in Fig. 3.30, the tip radius was reduced from 67±3 to 32 nm by growing a 252 nm oxide at 950°C. At 1100°C, the tip radius was reduced to only 50 nm when a 235 nm oxide was grown. The results reveal that the sharpening of emitter tips by thermal oxidation is not as efficient as plasma oxidation, especially when a higher thermal oxidation temperature is used.

The sharpening of emitter tips by oxidation is related to the stress at the Si-SiO_2 interface on nonplanar surfaces.[32,33] This stress is due to the increase in molar volume when Si is oxidized. At the regions of high curvature, such as the emitter tip apex, the stress retards the oxidation rate locally. It has been proposed that the viscous stress associated with the nonuniform deformation of oxide at nonplanar surfaces causes retardation of oxide growth at the regions of high curvature. This reduction in oxidation rate at the tip apex will result in a sharper tip when the grown oxide is removed.

The viscous stress, which is responsible for tip sharpening, is proportional to the oxide viscosity.[33] Higher oxide viscosity will increase the stress and therefore the tip sharpness. Since the oxide viscosity is higher at low temperatures, sharper emitter tips were developed at 950°C than at 1100°C thermal oxidation. At higher temperatures, the decreased viscosity will allow the oxide to flow, which relieves the stress. Therefore, one way to increase the tip sharpness is to increase the stress by reducing the oxidation temperature. However, for thermal oxidation, the oxidation rate decreases drastically at lower temperatures. For example, it took 35 hr to grow 252 nm of dry thermal oxide at 950°C. With plasma oxidation, since reactive oxygen species were generated in the plasma, 257 nm of plasma oxide was grown in 8 hr with the stage set at room temperature. Therefore, oxidation in an O_2 plasma allows the temperature to be lowered to enhance the sharpening effect. It also provides a good control over the plasma oxide thickness for sharpening Si emitter tips.

3.4.3 High Density Sharp Si Field Emitter Arrays

The Si etch profile was controlled to fabricate field emitter tips with different geometries. The SiO_2 mask profile and the pressure of Cl_2 plasma have the strongest influence on the Si etch profile, and therefore were used to control the shape of emitters. Figure 3.32(a) shows an Si field emitter formed by a two-step etch process at different pressures. Using 250 W microwave and

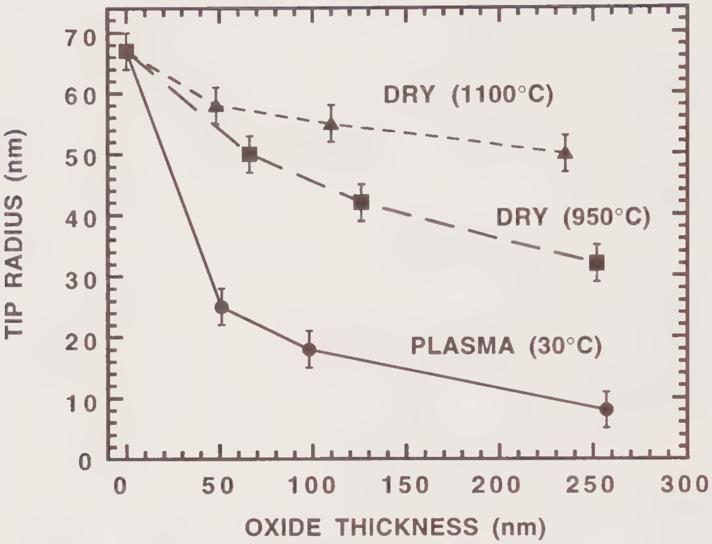


FIG. 3.30 Sharpening Si emitter tips by plasma oxidation. Oxygen plasma was generated by flowing 10 sccm of O_2 at 0.3 mTorr and applying 1000 W microwave power to the ECR source. The source distance was 8 cm and no rf power was applied to the stage. Thermal dry oxidation at 950 and 1100°C resulted in a larger tip radius compared to plasma oxidation. Error bars indicate the variations in the tip radius for different emitters.

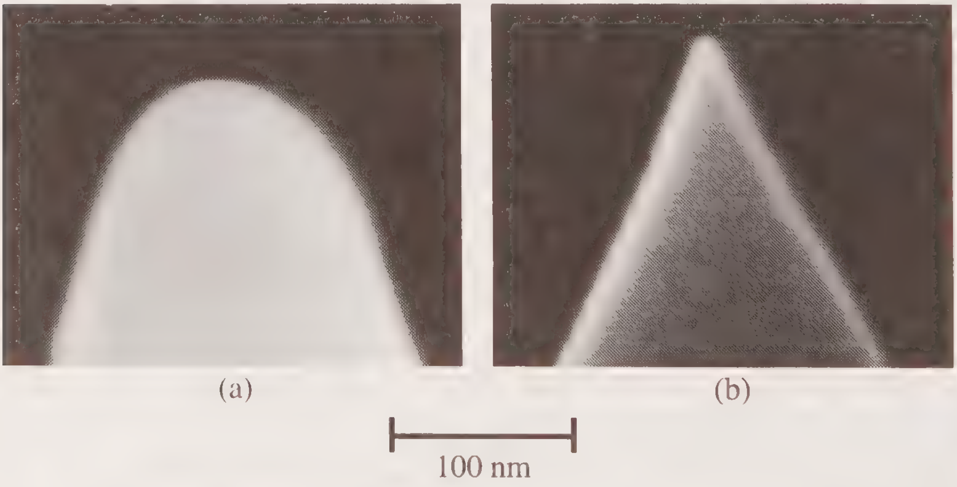
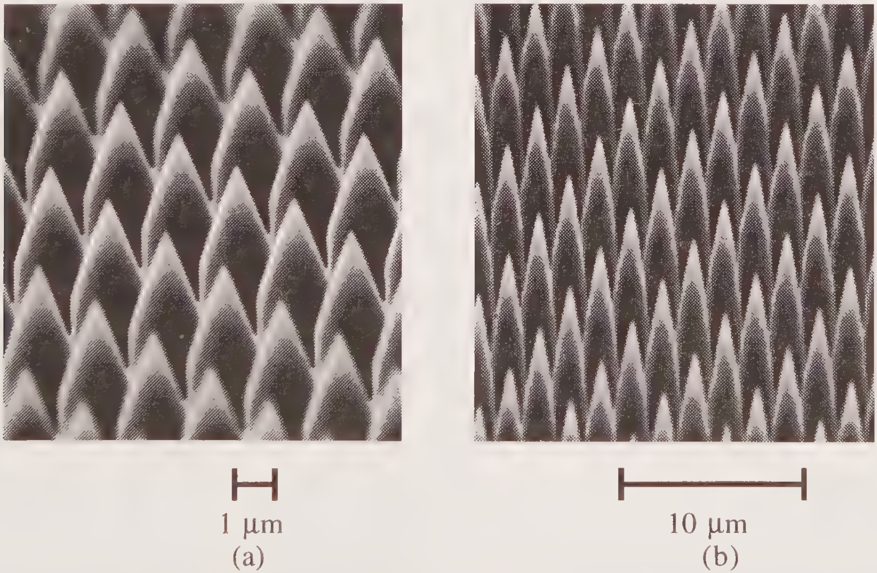


FIG. 3.31 Scanning electron micrographs of Si emitter tips sharpened by plasma oxidation. The oxidation conditions are the same as in FIG. 3.30. (a) The tip radius is 67 nm before oxidation. After growing a 257 nm plasma oxide and removing the oxide in buffered HF, the tip radius decreased to 8 nm in (b).

400 W rf power with 20 sccm of Cl_2, samples were first etched at 5 mTorr for 6 min. Then the pressure was decreased to 0.5 mTorr and samples were etched for an additional 11 min. At 5 mTorr, the lateral etch rate of SiO_2 was low and the lower part of the emitter with a sidewall angle of 85° was developed. As pressure was reduced to 0.5 mTorr, the lateral etch rate of SiO_2 increased and the top part of the emitter formed a cone shaped tip with a sidewall angle of 70°. According to the simulations, this Si emitter profile with a higher sidewall angle at the base provides better emission characteristics. An array of field emitter tips etched with 250 W microwave and 200 W rf power at 10 mTorr is also shown in Fig. 3.32(b). Due to the high pressure and low lateral etch rate of the SiO_2 mask, tall emitters with sidewall angles of 80° are formed. They are 11 μm in height, 2.2 μm in basewidth, and have a packing density of $4X10^6$ tip/cm^2. These results demonstrate that Si field emitters with high packing density can be generated by dry etching in a Cl_2 plasma. By controlling the etch mask profile and etch conditions in the plasma, Si emitters are fabricated with the desired sidewall angle, sharpness, and height to provide high emission current.



(a) (b)

FIG. 3.32 Scanning electron micrographs of field emitter tips and arrays. For (a), microwave power was 250 W and rf power was 400 W. The pressure was changed from 5 to 0.5 mTorr in a two-step process, resulting in emitter tips with an upper sidewall angle of 70°. (b) Sharp emitter tips with θ = 80° were fabricated using 250 W microwave power and 200 W rf power at 10 mTorr.

Dry etching and low temperature oxidation have been used to sharpen Si field emitter tips. Si emitters were formed by dry etching in a Cl_2 plasma and a mask erosion technique was used to control the emitter tip geometry. A typical tip radius for these Si emitters was 67 nm right after the oxide mask was completely eroded. Once the emitters were formed, further etching in the Cl_2 plasma could sharpen the emitter tips. The tip radius was reduced to 22 nm after 7 min of dry etching at 0.7 mTorr. Optical emission spectroscopy and mass spectrometry were used for real-time monitoring of the emitter sharpening. The optical emission signal of Si at 288.1 nm and the mass spectrometric signal for $^{63}SiCl^+$ showed sharp increases when the SiO_2 mask was completely removed and Si emitter tips were formed. Using these monitored signals, the etching can be stopped when the desired tip radius for the field emitters is reached.

Low temperature plasma oxidation was also used to sharpen the emitter tips. The tip radius decreased from 67 nm to 25 nm after growing 51 nm of plasma oxide followed by removal in buffered HF. When 257 nm of plasma oxide was grown, the tip radius was reduced to 8 nm. Thermal oxidation was also used to sharpen the emitter tips, but was not as effective as plasma oxidation. The tip radius only decreased from 67 nm to 50 nm after 235 nm of dry oxide was grown at 1100°C and removed in buffered HF. Using this dry etching technique, arrays of sharp emitter tips in Si with a 2.2 μm wide basewidth and 11 μm tall were fabricated and packing densities up to $1X10^7$ tip/cm^2 were achieved.

3.5 FABRICATION OF MIRRORS AND LASERS BY DRY ETCHING

Advances in Si micromachining technologies make it feasible to implement integrated micro-optical components.[34-36] Potential applications for these 3-D micro-optical components include switching, coupling/decoupling, interconnections for faster data transfer, as well as deflection/projection for high resolution image systems. Compared to conventional discrete optical components, these micromachined optical components are easier to integrate and package, simpler to align, and cost less to fabricate. Among all the micro-optical components, micromirrors and microlenses are two of the most important devices for manipulating and splitting optical beams.

Smooth sidewall morphology and vertical profile are necessary to minimize light scattering and maintain a collimated light beam when an etched sidewall is used as a mirror surface. Crystal orientation dependent wet etching in (110) Si has been used to fabricate high aspect ratio Si mirrors with good optical characteristics. However, the layout of microstructures is strictly limited and precise alignment along crystal orientation is needed. On the other hand, since dry etching does not depend on crystal orientation, more design flexibility and better device performance are expected.

A Cl_2 plasma was generated using an ECR source to etch high aspect ratio Si mirrors and lenses. Patterning of Ni etch masks by trilayer resist was used to minimize the roughness along the edges of the Ni mask. The effects of etch conditions on sidewall smoothness of etched mirrors were studied. In addition, by applying the deep etch-shallow diffusion technology, 40 μm tall Si micromirrors and microlenses were fabricated. These high aspect ratio Si micromirrors showed a vertical profile and smooth mirror surfaces, which are essential characteristics for high quality optical components.

3.5.1 Si Mirrors and Modulators

To fabricate high optical quality micromirrors in Si, a vertical profile and smooth sidewalls are essential. In optical lithography, rough edges are often observed due to interference fringes formed in the photoresist. Therefore, photoresist patterning and dry etching conditions were optimized to provide smooth sidewalls in the etched Si mirrors.

The Si mirrors were etched using a Cl_2 plasma generated in an ECR source with 100 W microwave power and 100 W rf power at 3 mTorr for 2 hr to a depth of ~20 μm. With the top imaging resist post-baked at 160°C for 10 min to smooth out the interference fringes, smoother sidewalls along the etched Si mirrors were obtained. Without postbake for the top layer photoresist, the measured root mean square (rms) roughness by atomic force microscopy (AFM) on the Si sidewall was 29.95 nm. Baking the imaging photoresist at 160°C for 10 min reduced the rms roughness to 18.69 nm. This suggests that the trilayer resist scheme with postbake can be used to significantly reduce the sidewall roughness on the etched Si mirrors.

After Si micromirrors were formed by dry etch in the ECR source, the roughness along the sidewalls can be further reduced by thermally oxidizing the Si microstructures followed by removal of SiO_2. Because Si oxidation rates are influenced by feature geometry and growth temperatures, roughness along Si sidewalls can be smoothed out with optimized oxidation conditions.

Figure 3.33 shows the Si sidewall rms roughness obtained from AFM as a function of oxidation time. Twenty micrometer tall Si micromirrors were formed by dry etching. The etched mirrors were oxidized using 2.7 slpm H_2 and 1.35 slpm O_2 wet oxidation at 1100°C. Before the oxidation, AFM results showed an rms roughness of 29.95 nm on the Si sidewalls. As the oxidation time was increased from 3.2 to 135 min, the oxide thickness on the planar (100) Si surface increased from 101 to 986 nm. The grown oxide was etched away by buffered HF with an etch rate of 105 nm/min. After the oxide was removed, the Si sidewall roughness decreased to only 5.93 nm after 135 min of oxidation. Since the amount of available oxidizing species depends on surface geometry, thermal oxidation rates are different for concave and convex surfaces. Typically oxide grows faster on convex surfaces than on concave ones. Ripples along Si sidewalls can be considered to consist of continuous concave and convex surfaces. The different oxidation rates on convex vs. concave surfaces result in smoother morphology after the SiO_2 is removed.

As shown in Fig. 3.34, higher oxidation temperature is more effective in smoothing the Si sidewalls. Wet oxidation was carried out at 900, 1000, and 1100°C for 220, 65, and 22 min to obtain a similar oxide thickness of 366, 371, and 349 nm, respectively. Even though the oxide thickness is similar, the Si sidewall oxidized at 900°C had a larger sidewall rms roughness of 17.12 nm, while those oxidized at 1100°C showed an rms sidewall roughness of

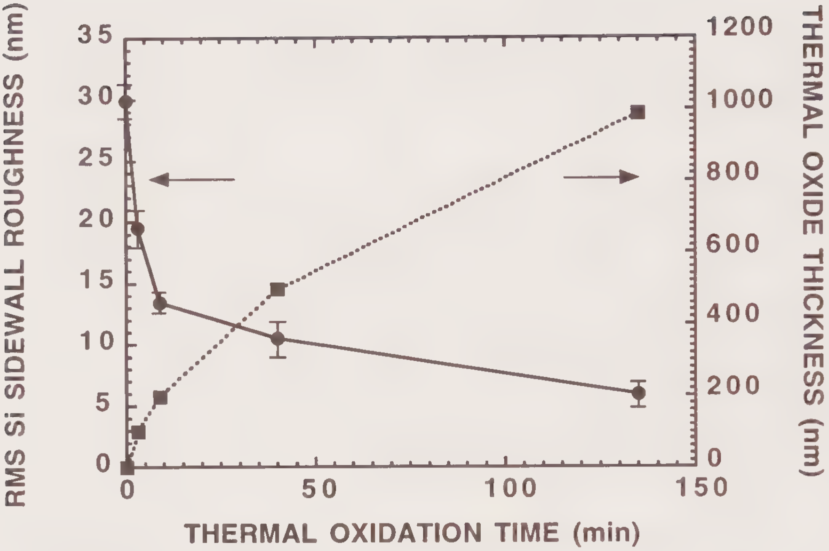


FIG. 3.33 Dependence of etched Si sidewall rms roughness on oxidation time. Si samples were etched to 20 μm deep using a Cl_2 plasma generated at 3 mTorr, 100 W microwave power, and 100 W rf power. The wet oxidation was carried out at 1100°C with 2.7 slpm of H_2 and 1.35 slpm of O_2.

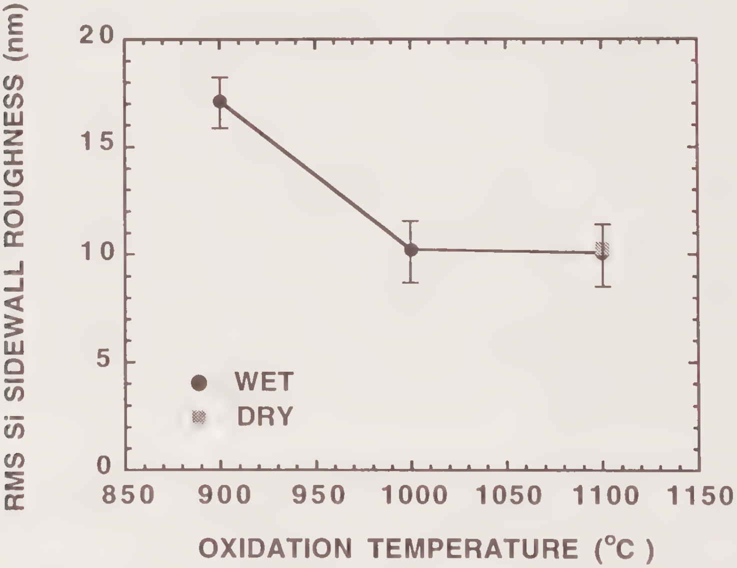
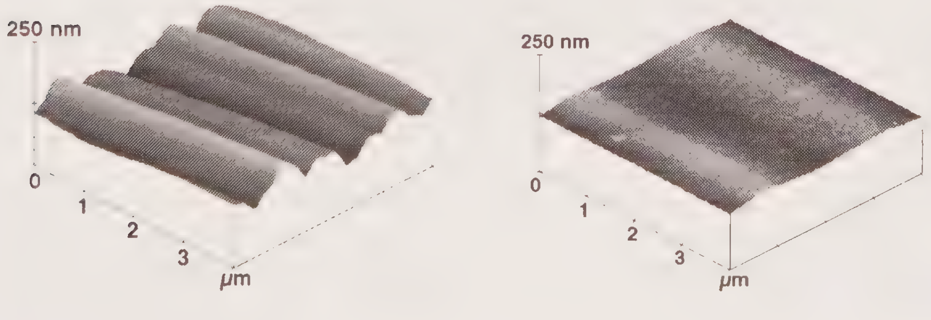


FIG. 3.34 Effects of oxidation temperature on Si sidewall roughness. Wet oxidation temperatures were 900, 1000, and 1100°C for 220, 65, and 22 min, respectively. Dry oxidation was done at 1100°C for 400 min.

10.07 nm. The dependence of sidewall smoothness on oxidation temperature could be related to the temperature dependent viscous stress for two dimensional oxide growth.[37] Locally enhanced compressive stress decreases the oxidation rates on concave or convex corners. At oxidation temperatures <965°C, the oxide viscous flow temperature, the stress cannot be relieved and the retardation of oxide growth is more severe. At higher oxidation temperatures, oxide viscosity is reduced and stress is also lower because oxide starts to flow, resulting in a higher oxidation rate and a smoother $Si\text{-}SiO_2$ interface. Therefore, a smoother sidewall could be obtained after oxide removal. In addition, dry oxidation at 1100°C for 400 min was also carried out for comparison. The oxide thickness obtained on the planar (100) Si surface was 349 nm. The Si sidewall rms roughness was 10.29 nm, which is essentially the same as the results obtained from wet oxidation.

High aspect ratio, released micromirrors and microlenses in Si were fabricated by applying the deep etch-shallow diffusion process. In this process, deep Si microstructures were etched with the ECR source, followed by a short B diffusion and a second dry etch. The Si microstructures were then bonded to glass and released by selective wet etch. As shown in the previous section, Si sidewall roughness could be reduced by thermal oxidation. During the B diffusion step at 1175°C, O_2 was introduced to minimize pitting of the Si surface and therefore Si was oxidized. After this heavily doped oxide layer was removed, Si sidewall smoothness was measured by AFM. The results were compared to the etched Si sidewall with and without wet oxidation at 1100°C for 22 min. After dry etching, the Si sidewall had an rms roughness of 29.95 nm, as shown in Fig. 3.35(a). Figure 3.35(b) shows the smoother sidewall after 2 hr of B diffusion and 2 hr of EDP etch at 110°C. These results suggest that smooth Si sidewalls can be inherently obtained in the deep etch-shallow diffusion process used to fabricate high aspect ratio, released Si microstructures because of the B diffusion and selective wet etch steps.



(a) (b)

FIG. 3.35 Scanned sidewall surface morphology using AFM. (a) Si microstructure etched using the same conditions as in FIG. 3.33 and (b) after B diffusion and EDP etch for 2 hr at 110°C.

Released Si micromirrors were fabricated by the deep etch-shallow diffusion process. The micromirrors were etched down to 40 µm deep and diffused with B for 6.5 hr. After an unmasked second dry etch to remove the heavily doped layer near the bottom of the trenches, these micromirrors were bonded to Corning #7740 glass and released in EDP for 4 hr. Figure 3.36(a) shows a released, 40 µm tall, 2 µm wide micromirror with both sides anchored on the glass substrate. The deep grooves in glass, formed by cutting halfway through the glass wafer, were designed to increase the clearance to minimize light scattering from the glass substrate. With the smooth sidewall and vertical profile, these micromirrors are expected to have high reflectivity.

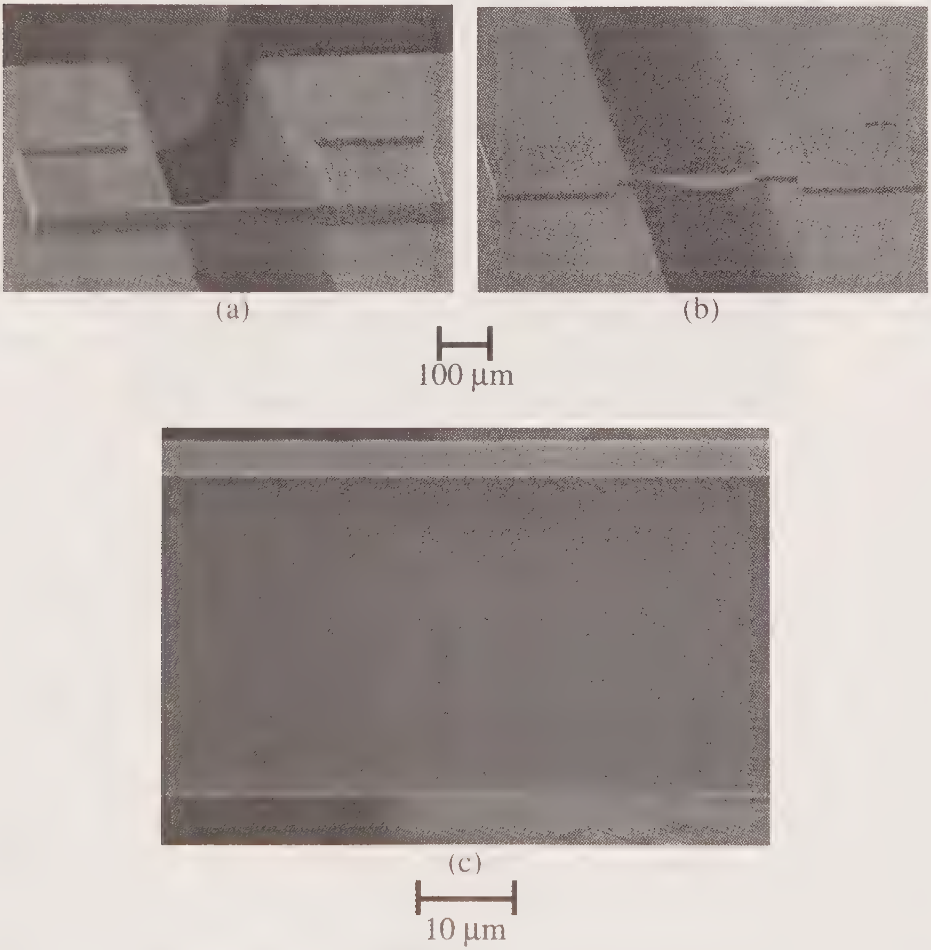
A potential problem often observed in optical systems is the coupling loss between the optical components and the fibers. This originates from the scattering of light as it passes through rough fiber ends. Most of the light is no longer collimated after refraction due to rough fiber end surfaces and only a small percentage of the light can be collected by other components with finite apertures. To improve the coupling efficiency, optical lenses have been widely used. However, Si microlenses have not been demonstrated. Because of the curvature, these microlenses cannot be implemented by wet etching. Therefore, high aspect ratio dry etching of Si is required for high performance microlenses. Figure 3.36(b) shows an Si microlens that was 40 µm tall with radius of curvature of 50 µm. The microlens was fabricated by the deep etch-shallow diffusion process discussed earlier. These micromirrors and microlenses have a smooth sidewall surface and vertical profile, as shown by the close-up view of the mirror surface in Fig. 3.36(c). These surfaces are expected to provide high optical quality. The micromachined mirrors and lenses can be integrated to form passive optical components such as a beam splitter, coupler, or modulator. They can also be combined with other active optical components for sensor applications.

3.5.2 GaAs Mirrors and Waveguides

Etched mirrors and waveguides in compound semiconductors have also been fabricated. These are critical components for optoelectronic circuit integration and for fabricating surface-emitting laser arrays and diode ring lasers.[38,39] Dry etching allows these mirrors to be fabricated with high packing density and vertical profiles. In order to achieve high reflectivity on these mirrors, the surfaces of the sidewalls should be smooth because surface roughness can cause light scattering and reduce reflectivity. In addition, etch-induced defects have to be minimized because these defects can act as nonradiative recombination centers. A vertical etch profile is also crucial to ensure that all the light will be reflected with the same direction.

Dry etched mirrors for the $In_{0.20}Ga_{0.80}As/GaAs$ based waveguides were fabricated using a Cl_2/Ar plasma generated with an ECR source.[40] The epitaxial layers consisted of an $Al_{0.30}Ga_{0.70}As$ cladding layer and $In_{0.20}Ga_{0.80}As/GaAs/In_{0.20}Ga_{0.80}As$ multiple quantum wells (MQW). Reflectivity was measured on the totally reflecting mirrors at the corner of a

90° bent waveguide. The width of the waveguides was 5 μm. Since the corner mirror is inclined at 45° to the waveguides, total internal reflection occurs. The mirror reflectivity measurements were made by coupling light from a 1.15 μm HeNe laser into the cleaved facet of the L-shaped waveguides, and the light output from the other end was measured using a Ge detector. The reflectivity of the mirrors is given by the ratio of the light output from the L-shaped waveguide to that from a straight waveguide with equivalent length. A 125 nm thick Ni was lifted off as the etch mask. This Ni mask can be removed by HCl after dry etching.



(a) (b)

FIG. 3.36 Scanning electron micrographs of (a) 40 μm thick, 2 μm wide Si vertical micromirrors bonded onto glass and released in EDP. Deep grooves in glass were formed by cutting 300 μm deep into the glass with a dicing saw. (b) A Si microlens that is 40 μm deep with a radius of curvature of 50 μm. (c) A closeup view of the mirror surface showing the smooth sidewall and vertical profile.

Figure 3.37 shows a typical $In_{0.20}Ga_{0.80}As/GaAs$ mirror etched down to 2.8 μm deep. The mirrors were etched in a Cl_2/Ar plasma using 10% Cl_2

generated with 50 W microwave power and 70 W rf power at 0.5 mTorr. It can be seen that a vertical profile and smooth surface morphology have been obtained. The etch rate was 0.11 µm/min. The selectivity of GaAs to Ni was 42 under these etch conditions, so the 125 nm thick Ni mask is sufficient for this etch depth. This highly controllable profile and surface morphology are critical for high reflectivity mirrors. These etched mirrors were of very high optical quality, and reflectivities up to 93% have been achieved.

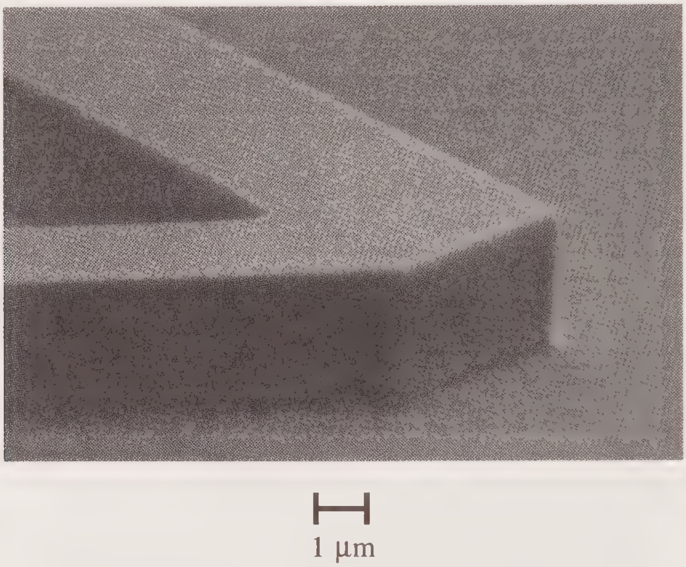


FIG. 3.37 Scanning electron micrograph of the $In_{0.20}Ga_{0.80}As/GaAs$ total internal reflecting mirror. The mirror was etched in a Cl_2/Ar plasma using 10% Cl_2, 50 W microwave power and 70 W rf power at 0.5 mTorr, to an etch depth of 2.8 µm.

This dry etching technique has also been applied to the fabrication of microcavities in $In_{0.20}Ga_{0.80}As/GaAs$. Significant improvement in the threshold current density has been predicted for microcavity structures in which the width of the optical cavity is reduced to 0.3 µm.[41] Figure 3.38 shows 0.25 µm wide microcavities etched to a depth of 2.8 µm with vertical profile and smooth surface. This result demonstrates that this dry etching technology is also suitable for fabricating high aspect ratio microcavities. In addition, dry etching of $In_{0.90}Ga_{0.10}As_{0.22}P_{0.78}/InP$ mirrors has also been achieved with vertical profile and smooth sidewalls.

The reflectivity for mirrors etched under various conditions was measured and compared so that the optimal conditions for fabricating high quality mirrors could be found. Figure 3.39 shows the changes in the mirror reflectivity as a function of rf power. The sidewall damage depths extracted from GaAs conducting wires etched at different rf power are also shown for comparison.[42] The samples were etched with 10% Cl_2 in a Cl_2/Ar plasma

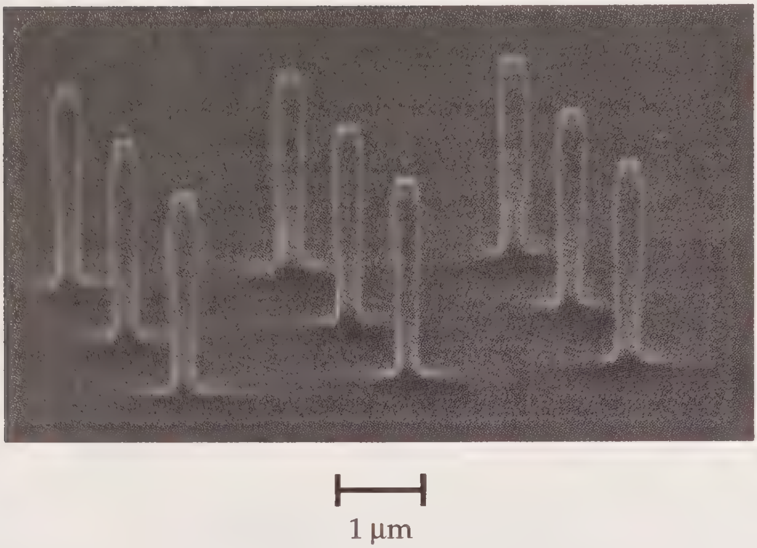


FIG. 3.38 0.25 μm wide microcavities in $In_{0.20}Ga_{0.80}As/GaAs$ etched down to 2.8 μm. The etch conditions are the same as shown in FIG. 3.37.

generated at 0.5 mTorr with 50 W microwave power and rf power ranging from 70 to 200 W. A vertical profile and smooth sidewalls were achieved on all the mirrors etched within this rf power range. The etch depth was 2.8 μm. It can be seen that even though the sidewall damage depth increases substantially from 9.9 to 20.4 nm with rf power, there is no significant change in the mirror reflectivity. The reflectivity is 84% at 70 W rf power and 93% at 200 W rf power. Since this is within the range of measurement uncertainty, reflectivity is mostly independent of the rf power used during etching. These results suggest that the additional electrical degradation on the sidewalls induced at higher rf power does not cause any reduction in the mirror's reflectivity.

Another crucial parameter in dry etching is the ion flux. Similar to the case of the ion energy, there is no significant change in the reflectivity with ion flux. The reflectivity remains at ~90% when various microwave powers are used for mirror etching. However, when a higher Cl_2 concentration was used during etching, the profile of the etched mirrors was undercut, the sidewalls were rougher, and a significant decrease in the reflectivity down to 47% was observed. These results suggest that reflectivity for dry etched mirrors is most sensitive to the etch profile and sidewall roughness and not so much to the ion energy or the ion flux. Triangular ring lasers with high quality etched mirrors were fabricated, and threshold current as low as 3.7 mA was achieved. A vertical profile with smooth sidewalls was also demonstrated on $Ga_{0.10}In_{0.90}As_{0.22}P_{0.78}/InP$ mirrors down to an etch depth of 6.5 μm using a Cl_2/Ar plasma.

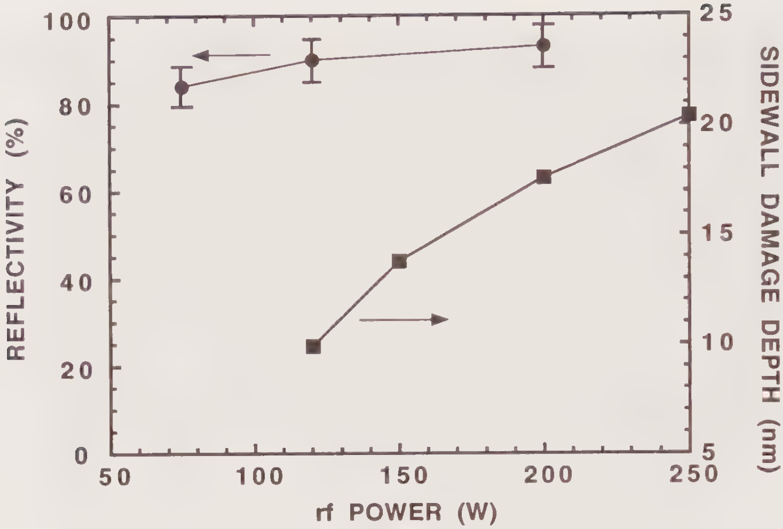


FIG. 3.39 Reflectivity (●) and sidewall damage depth (■) as a function of rf power. The samples were etched under conditions similar to those shown in FIG. 3.37 with rf power ranging from 70 to 200 W.

3.5.3 Horizontal Distributed Bragg Reflector Mirrors

For full integration of waveguide lasers into optoelectronic circuits, it is required that an alternative to cleaved mirrors be employed. Mirrors can be formed by etching smooth, vertical facets or by modifying the material structure so that a distributed Bragg reflector (DBR) is formed. Etched mirrors typically have low reflectivity and require the additional steps of dielectric deposition or passivation to improve reflectivity. DBR mirrors fabricated by etch and regrowth[43] rely on a very small change in the index of refraction, *n*, and several hundred periods are required to obtain high reflectivity. The passive mirror region can be quite long for these structures.

A large change in the refractive index can be achieved by etching a periodic structure consisting of the semiconductor material and an air gap. There are several advantages to using horizontal etched DBR mirrors to define the cavity of a waveguide laser. First, due to a large difference in refractive index between air and semiconductor materials, few periods are needed to obtain high reflectivities. Therefore, the throughput for patterning horizontal DBR mirrors can be high and the lateral dimension of the passive region can be minimized. Second, since the reflectivity is dependent upon the number of periods, mirror reflectivity can be high at one end of the laser and slightly lower at the other end to optimize the output intensity. In addition, the device is integrable with other on-chip optical devices. The process is not complex since it only requires one electron beam lithography step after all the layers have been grown, and one dry etch step to define the mirrors and the waveguide. The patterning and etching of these mirrors for both InP and GaAs- based lasers have been investigated.

Horizontal DBR mirrors in InP and GaAs have been developed for the current optical systems with optimized lithography and etching. For maximum reflectivity at 1.55 and 0.85 μm, the semiconductor should be 122 and 59 nm wide for λ/4 period mirrors, and the air gap should be 388 and 213 nm wide, respectively.[44] The required electron beam doses for different linewidths were measured after liftoff using scanning electron microscopy. Interpolation was used to obtain the appropriate dose needed to provide 122 nm wide lines on InP and 59 nm wide lines on GaAs. Figure 3.40 shows a three period DBR stack adjacent to a waveguide region on InP. Both features were written using electron-beam lithography. Theoretically, this mirror structure should have a reflectivity of 99% for a planar optical wave. The dose of the lines farthest from the waveguide was increased to correct for the proximity effect. A dose of 6.1 nC/cm is required to obtain the linewidth of 122 nm on InP, and a dose of 2.3 nC/cm yields 59 nm Ni lines on GaAs.

FIG. 3.40 Top view of waveguide and distributed Bragg reflector mirror structure patterned on InP using electron beam lithography, evaporated Ti/Ni, and liftoff. The Ti/Ni lines are 122 nm and the spacing between lines is 388 nm.

The two most important issues for the etching of the DBR mirrors are etch selectivity to the masking material and the etch profile. The etch should remove the upper cladding region, the active region, and part of the lower cladding layer. For a typical InP-based device structure,[45] this requires an etch over 2.1 μm deep, while for a GaAs/AlGaAs waveguide laser, an etch of 1.4 μm is sufficient. Most of the InP-based device in terms of layer thickness is the InP cladding layer. For the GaAs/AlGaAs laser, the cladding is AlGaAs, but the etch characteristics of GaAs are similar to those of AlGaAs. Therefore, InP and GaAs substrates have been used to optimize lithography and etch conditions.

The high resolution patterning requires use of a thin polymethylmethacrylate (PMMA) layer and limits the mask layer thickness. For a 100 nm mask, a selectivity >34 is required for InP. The ECR source allows the ion energy to be controlled by adjusting the rf power applied to the stage. Shown in Fig. 3.41 is the effect of increasing the rf power on the InP and Ni etch rates. The etch conditions were 250 W of microwave power, Cl_2/Ar flow of 3/17 sccm, 1.5 mTorr chamber pressure, and a stage temperature of 350°C. The self induced dc bias changed from 130 to 260 V as rf power was increased from 100 to 250 W. At an rf power of 100 W, the InP etch rate is 879 nm/min and the Ni sputter rate is 6 nm/min. This yields a selectivity of 147. As the rf power increases, both etch rates increase. At 250 W rf power, the InP etch rate is 1.11 μm/min and the Ni sputter rate is 20 nm/min. However, this leads to a decrease in the selectivity of InP to Ni to 55. To etch the deep mirror structures, high selectivity must be maintained. At the low rf power of 100 W, where the selectivity was largest, the etched surface is somewhat rough due to inefficient removal of the $InCl_x$ etch products, leading to preferential etching and roughening of the surface. By increasing the ion energy, the physical component of the etch is increased. This tends to equalize the etch rates of the In and P compounds and a smooth surface is obtained, but the etch selectivity is reduced. Therefore, a tradeoff exists between how deep a mirror can be etched and how smooth the surface can be.

A λ/4 period mirror in InP etched at 200 W rf power and 350°C is shown in Fig. 3.42(a). The gaps are etched 2.0 μm deep and the region outside the waveguide is etched 3.4 μm deep. The gap is etched through the active region into the lower cladding while the area outside the mirror stack is etched through the lower cladding. The reflectivity of the mirror should not be degraded since the gap is etched into the lower cladding. It can be seen that a slower etch rate and a rough surface occur near the waveguide and in between the mirror plates. These suggest that the roughness is due to the redeposition of sputtered etch mask onto the surface, causing micromasking. One method to reduce the surface roughness is to increase the gap spacing. This allows more ions to enter the gap and assist in removal of etch products and redeposited materials. The mirrors can be designed as a 3λ/4, 5λ/4 or 7λ/4 period grating structure. A 3λ/4 period mirror in InP is shown in Fig. 3.42(b). The reflectivity of the mirror will decrease slightly but the operation of the mirror is the same.

Figure 3.43 shows how the microloading effect is reduced as the gap between mirrors is increased from 388 nm (λ/4) to 2.72 μm (7λ/4). At 1 mTorr, the normalized etch rate (obtained by dividing the etch rate of the gap by the etch rate obtained for an open area) increases from 53 to 100% as the gap spacing is increased. In addition, Fig. 3.43 shows that the microloading effect can be reduced by decreasing the pressure. For the 3λ/4 linewidth, at 0.7 mTorr the

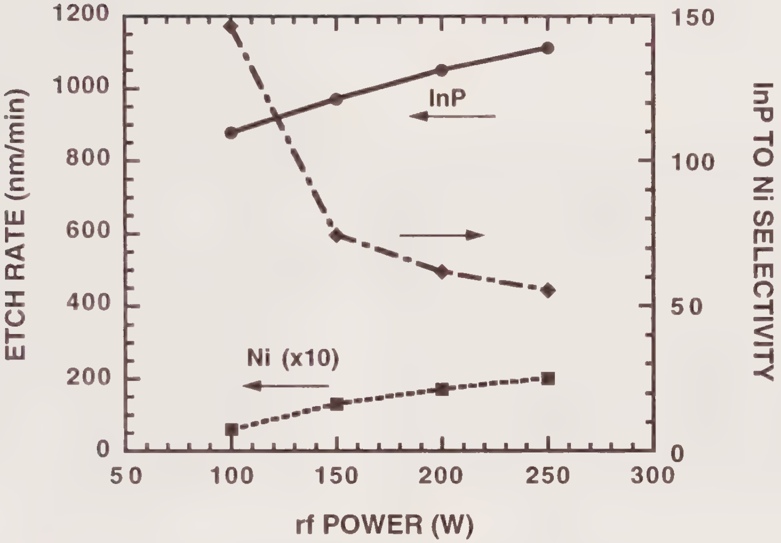


FIG. 3.41 Etch rates of InP and Ni increase with increasing rf power, and the selectivity of InP to Ni decreases. The etch conditions were 250 W microwave power, Cl_2/Ar at 3/17 sccm, 1.5 mTorr pressure, an 8 cm source distance, and a stage temperature of 350°C.

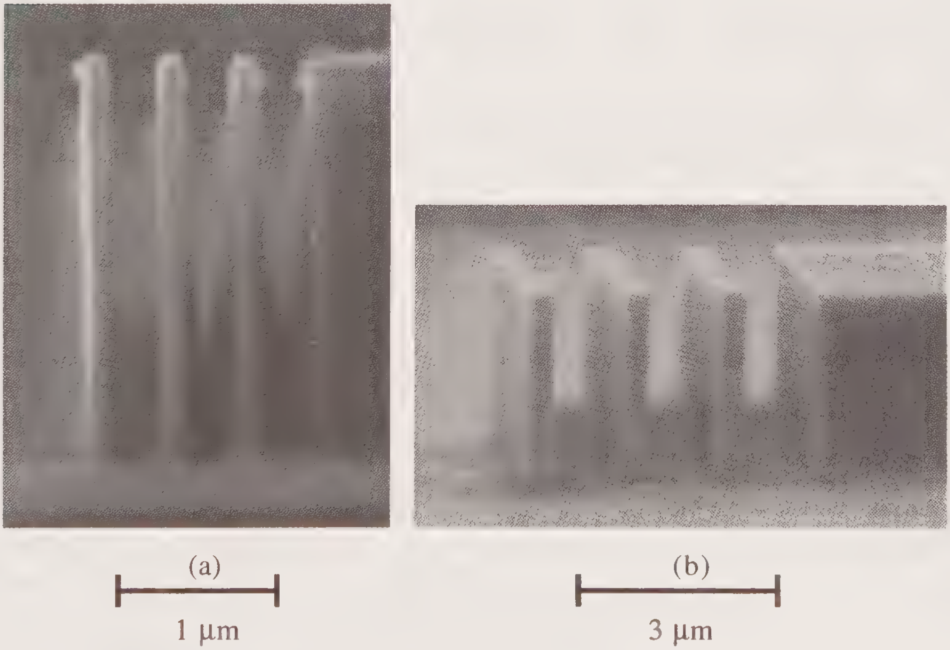


FIG. 3.42 DBR mirrors etched in InP with (a) λ/4 period and (b) 3λ/4 period. The etch conditions were 250 W microwave power, 200 W rf power, Cl_2/Ar at 4/16 sccm, 1.0 mTorr pressure, an 8 cm source distance, and a stage temperature of 350°C.

normalized etch rate is 96%, while at 1.3 mTorr, the normalized etch rate is only 76%.

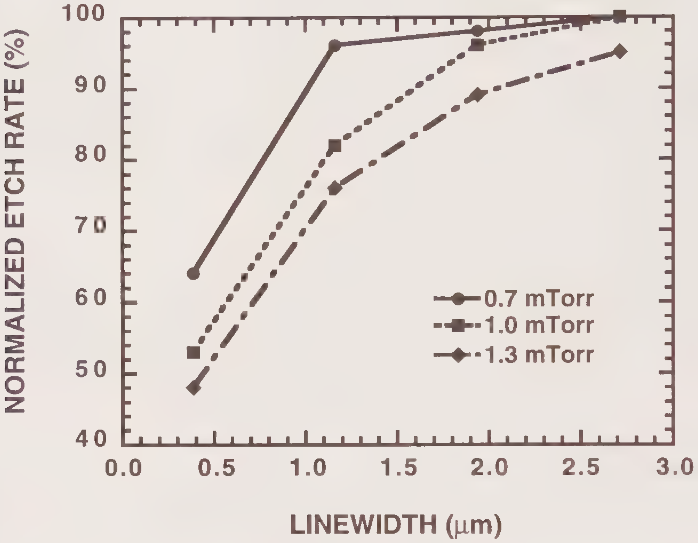


FIG. 3.43 Effects of linewidth and pressure on the etch rate of InP for various gap widths. The etch conditions were 250 W microwave power, 200 W rf power, Cl_2/Ar at 2/18 sccm, an 8 cm source distance, and 350°C stage temperature.

High selectivity and a vertical profile must also be obtained for the GaAs-based DBR mirror. The rf power was found to have a large effect on the selectivity of GaAs to Ni. For 50 W rf power, the GaAs etch rate is 287 nm/min, the Ni sputter rate is 2.8 nm/min, and the selectivity is 102. At 200 W rf power, the selectivity decreases to 46, due to an increase in the Ni sputter rate to 8.6 nm/min. The 3X increase in the Ni sputter rate is caused by the increasing ion energy with rf power. At 50 W rf power, $|V_{dc}|$ is only 40 V, while at 200 W rf power, $|V_{dc}|$ is 255 V. The etch conditions were 50 W microwave power, Cl_2/Ar flow at 4/16 sccm, 1 mTorr pressure, 8 cm source distance, and 25°C stage temperature. For etching the mirrors, the Cl_2 percentage was kept at 20% and the profile remained vertical. Figure 3.44 shows a 1.5 μm deep, λ/4 mirror etched with 20% Cl_2 and 100 W rf power. The etch depth of the 213 nm wide gap is 64% of the open area. The 3λ/4 mirror that had 638 nm wide gaps had a normalized etch rate of 75%, and the etch depth was 1.8 μm. As in the case of InP mirror stacks, the gaps between mirrors would be etched into the lower cladding and the region outside the waveguide would be etched through the lower cladding.

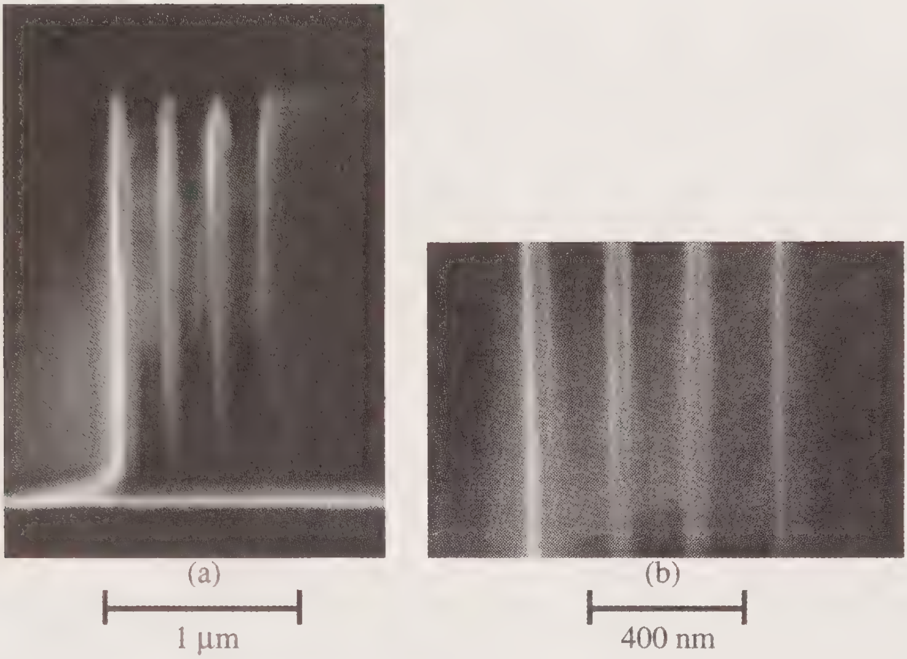


FIG. 3.44 Three period distributed Bragg reflector mirror structure etched in GaAs. The etch conditions were 50 W microwave power, 100 W rf power, Cl_2/Ar gas flow at 4/16 sccm, 1 mTorr pressure, and an 8 cm source distance.

Narrow lines have been patterned on InP and GaAs using electron beam lithography for the fabrication of horizontal DBR mirrors. The Ni lines deposited on InP were 122 nm wide and on GaAs were 59 nm wide for $\lambda/4$ DBR mirrors. This should allow high reflectivity mirrors to be fabricated for lasers emitting at 1.55 and 0.85 µm. The etched mirrors allow for fabrication of optoelectronic integrated circuits with increased design flexibility. The rf power, Cl_2 percentage in Ar, pressure, and temperature were studied to optimize the etch selectivity and etch profile. Use of low rf power improved the selectivity, while use of low Cl_2 percentage and low pressure allowed vertical profiles to be etched. High temperature was used to improve the InP to Ni selectivity from 39 to 62. The increase in the In emission intensity is used to signify the beginning of a high rate etching of InP, and at a high stage temperature it increased ~1 min earlier than when the stage was at room temperature. Two micrometer tall InP and 1.5 µm tall GaAs, $\lambda/4$ horizontal DBR mirrors were etched. These etch depths are sufficient to clear the top cladding layer and the active region of most devices, and should provide high reflectivity. The effect of gap spacing on etch rate was studied. Use of low pressure allowed the $n\lambda/4$ gratings to be etched at nearly the same rate as the open areas, allowing more of the lower cladding to be removed.

3.6 SUMMARY

High aspect ratio etching of polyimide in an O_2 plasma generated by an ECR source has been demonstrated. Etch rate, profile, and selectivity were found to depend on microwave power, rf power, pressure, etch time, and gas composition. Etch rate was found to increase with microwave and rf power. Etch anisotropy increases with rf power but decreases with pressure and etch time. With 750 W microwave power, 300 W rf power, and 0.5 mTorr pressure, the polyimide etch rate was 0.91 µm/min and the anisotropy was 0.92. Etch selectivity between polyimide and the Ti mask was 3150:1 at 50 W and 536:1 at 300 W rf power. Compared with conventional reactive ion etching, etching using an ECR source provides an etch rate that is typically 10 times faster and etch selectivity that is 4-6 times higher. High aspect ratio (>15:1) microstructures in polyimide that were 50 µm deep and 3 µm wide have been obtained using an O_2 plasma generated by an ECR source.

A novel etch-diffusion process has been developed to fabricate high aspect ratio microstructures in Si with a short diffusion time. Anisotropic etching of Si using a Cl_2 plasma generated by an ECR source is used to fabricate resonators with a vertical etch profile, smooth surface morphology, and high selectivity over an Ni etch mask. A short B diffusion time of 5.5 hr at 1175°C is sufficient for 40 µm thick, 12 µm wide Si microstructures. Using this etch-diffusion process, released Si microstructures up to 40 µm thick with 2 µm gaps have been demonstrated by deep Si etching and shallow B diffusion. For the 35-µm-thick laterally driven resonators, the resonant frequency was 2.49 kHz and Q was 23.

A mask erosion technique was implemented to fabricate field emitter tips and arrays. SiO_2 was used as an erosive mask to etch Si. A Cl_2 plasma generated by an ECR source was used to etch the Si emitters. Among other factors, the SiO_2 mask profile and pressure had the largest effects on the Si etch profile, while microwave and rf power were used to adjust the etch rates without changing the Si profile significantly. Once the emitters were formed, further etching in the Cl_2 plasma could sharpen the emitter tips. The tip radius was reduced to 22 nm after 7 min of dry etching at 0.7 mTorr. Low temperature plasma oxidation can also be used to sharpen the emitter tips down to 8 nm. Sharp emitter tips as high as 11 µm with a packing density of $4X10^6$ tip/cm^2 were fabricated.

Micromachined vertical mirrors and lenses in Si were fabricated as micro-optical components. The smoothness of the Si mirrors depends on the mask material, lithography, and dry etch conditions. By optimizing the conditions used to fabricate the mirrors, the sidewall roughness was only 18.69 nm compared to 98.99 nm when low rf power and a rough etch mask were used. In addition, thermal oxidation followed by oxide removal was able to reduce etched Si mirror roughness. After oxide removal, the sidewall roughness decreased to 5.93 nm after 135 min of wet oxidation at 1100°C. Vertical micromirrors that were 40 µm tall and 2 µm wide were bonded onto the glass substrate and released. Microlenses that were 40 µm thick with a radius of curvature of 50 µm have also been demonstrated.

High quality $In_{0.20}Ga_{0.80}As/GaAs$ total internal reflecting mirrors have been demonstrated. The etched mirrors have smooth sidewalls and a vertical profile. Mirror reflectivities up to 93% have been measured. It was found that the reflectivity of the etched mirrors was not sensitive to either the ion energy or the ion flux used for the etching. As long as a vertical profile and smooth surface morphology are maintained, the reflectivity remains at ~90%. However, significant degradation of the mirror reflectivity to only 47% was found on mirrors with a slight undercut profile and rougher sidewalls. Triangular ring lasers with etched mirrors have also been fabricated, and threshold current as low as 3.7 mA has been demonstrated, showing the high quality of the etched mirrors.

Horizontal DBR mirrors were formed in InP and GaAs for high reflectivity and high efficiency lasers emitting at 1.55 and 0.85 μm. The etched mirrors allow for fabrication of optoelectronic integrated circuits with increased design flexibility. Use of low rf power improved the etch selectivity, while use of low Cl_2 percentage and low pressure allowed vertical profiles to be etched. Two micrometer tall InP and 1.5 μm tall GaAs, $\lambda/4$ horizontal DBR mirrors were etched. These etch depths are sufficient to clear the top cladding layer and the active region of most devices, and should provide high reflectivity. The effect of gap spacing on etch rate was studied. Use of low pressure allowed the $n\lambda/4$ gratings to be etched at nearly the same rate as the open areas, allowing more of the lower cladding to be removed.

REFERENCES

1. O. Joubert, J. Pelletier, C. Fiori, and T. A. Nguyen Tan, "Surface Mechanisms in O_2 and SF_6 Microwave Plasma Etching of Polymers," *J. Appl. Phys.* **67**, pp. 4291-4296, 1990.
2. S. W. Pang, K. T. Sung, and K. K. Ko, "Etching of Photoresist Using Oxygen Plasma Generated by a Multipolar Electron Cyclotron Resonance Source," *J. Vac. Sci. Technol. B* **10**, pp. 1118-1123 1992.
3. W. H. Juan and S. W. Pang, "High Aspect Ratio Polyimide Etching Using an Oxygen Plasma Generated by Electron Cyclotron Resonance Source," *J. Vac. Sci. Technol. B* **12**, pp. 422-426, 1994.
4. E. W. Becker, W. Ehrheld, P. Hagmann, A. Maner, and D. Münchmeyer, "Fabrication of Microstructures with High Aspect Ratio and Great Structure Heights by Synchrotron Radiation Lithography, Galvanoforming, and Plastic Moulding (LIGA Process)," *Microelectron. Eng.* **4**, pp. 35-56, 1986.
5. G. Engelmann, O. Ehrmann, J. Simon, and H. Reichl, "Fabrication of High Depth-to-Width Aspect Ratio Microstructures," *Proc. IEEE Workshop on Micro Electro Mechanical Systems* (MEMS'92), Travemünde, Germany, 1992, pp. 93-98.
6. A. B. Frazier and M. G. Allen, "Metallic Microstructures Fabricated Using Photosensitive Polyimide Electroplating Molds," *IEEE J. Microelectromech. Syst.* **2**, pp. 87-94, 1993.

7. R. A. Gottscho, C. W. Jurgensen, and D. J. Vitkavage, "Microscopic Uniformity in Plasma Etching," *J. Vac. Sci. Technol. B* **10**, pp. 2133-2147, 1992.
8. W. H. Juan, S. W. Pang, A. Selvakumar, M. W. Putty, and K. Najafi, "Using Electron Cyclotron Resonance Source to Etch Polyimide Modes For Fabrication of Electroplated Microstructures," *Tech. Dig. Solid-State Sensors and Actuators Workshop*, pp. 82-85, 1994.
9. M. O'Donnell and L. J. Thomas, "Efficient Synthetic Aperture Imaging from a Circular Aperture with Possible Application to Catheter-based Imaging," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **39**, pp. 366-380, 1992.
10. J. A. Bride, S. Baskaran, N. Taylor, J. W. Halloran, W. H. Juan, S. W. Pang, and M. O'Donnell, "Photolithographic Micromolding of Ceramics Using Plasma Etched Polyimide Patterns," *Appl. Phys. Lett.* **63**, pp. 3379-3381, 1993.
11. Y. B. Gianchandani and K. Najafi, "A bulk silicon dissolved wafer process for microelectromechanical devices," *IEEE J. of Microelectromech. Syst.* **1**, pp. 77-85, 1992.
12. Z. L. Zhang and N. C. MacDonald, "An RIE process for submicron, Si electromechanical structures," in *Proc. Int. Conf. Solid-State Sensors and Actuators*, (Transducers'91), San Francisco, CA, June 1991, pp. 520-523.
13. C. Keller and M. Ferrari, "Milli-Scale Polysilicon Structures," in *Tech. Dig. Solid-State Sensors and Actuator Workshop*, Hilton Head, SC, June 1994, pp. 132-137.
14. W. H. Juan and S. W. Pang, "High Aspect Ratio Si Etching For Microsensor Fabrication," *J. Vac. Sci. Technol. A* **13**, pp. 834-838, 1995.
15. W. H. Juan and S. W. Pang, "Control of Si Etch Profile for Microsensor Fabrication," *J. Vac. Sci. Technol. A* **14**, pp. 1189-1193, 1996.
16. R. R. Burke, J. Pelletier, C. Pomot, and L. Vallier, "Distributed Electron Cyclotron Resonance in Silicon Processing: Epitaxy and Etching," *J. Vac. Sci. Technol.* A **8**, pp. 2931-2938, 1990.
17. J. W. Coburn and H. F. Winters, "Conductance Considerations in the Reactive Ion Etching of High Aspect Ratio Features," *Appl. Phys. Lett.* **55**, pp. 2730-2732, 1989.
18. J. Zheng, R. P. Brinkmann, and J. P. McVittie, "The Effect of the Presheath on the Ion Angular Distribution at the Wafer Surface," *J. Vac. Sci. Technol.* A **13**, pp. 859-864, 1995.
19. Y. B. Gianchandani and K. Najafi, "A bulk silicon dissolved wafer process for microelectromechanical devices," *IEEE J. of Microelectromech. Syst.* **1**, pp. 77-85, 1992.
20. W. Chu and M. Mehregany, "A study of residual stress distribution through the thickness of p^+ silicon films," *IEEE Trans. Electron Devices* **34**, pp. 1245-1250, 1993.
21. W. C. Tang, M. G. Lin, and R. T. Howe, "Electrostatic comb drive levitation and control method," *IEEE J. of Microelectromech. Syst.* **1**, pp. 170-178, 1992.

22. C. A. Spindt, C. E. Holland, I. Brodie, J. B. Mooney, and E. R. Westerberg, "Field-emitter arrays applied to vacuum fluorescent displays," *IEEE Trans. Electron Devices* **ED-36**, 225, 1989.
23. H. H. Busta, "Vacuum microelectronics-1992," *J. Micromech. Microeng.* **2**, 43, 1992.
24. H. C. Lee and R. S. Huang, "A study on field-emission array pressure sensors," *Sens. Actuat. A* **34**, 137, 1992.
25. H. F. Gray, C. T. Sune, and G. W. Jones, "Si field-emitter arrays for cathodoluminescent flat-panel displays," *J. Soc. Inf. Disp.*, 143, 1993.
26. C. A. Spindt, C. E. Holland, A. Rosengreen, and I. Brodie, "Field-emitter arrays for vacuum microelectronics," *IEEE Trans. Electron Devices* **ED-38**, 2355, 1991.
27. J. T. Trujillo and C. E. Hunt, "Fabrication of gated silicon field-emission cathodes for vacuum microelectronics and electron-beam applications," *J. Vac. Sci. Technol. B* **11**, 454, 1993.
28. M. Rakhshandehroo and S. W. Pang, "Dry Etching of Field Emitter Tips in Si," *J. Vac. Sci. Technol. B* **14**, pp. 612-616, 1996.
29. M. Rakhshandehroo, F. Sukardi, and S. W. Pang, "Simulation and Fabrication of High Density Si Field Emitters," *J. Vac. Sci. Technol. A* **14**, pp. 1832-1838, 1996.
30. B. Chapman, *Glow Discharge Processes*, John Wiley, New York, 1980, p 247.
31. K. T. Sung and S. W. Pang, "Low Temperature Silicon Oxidation with Oxygen Plasma Generated by an Electron Cyclotron Resonance Source," *J. Vac. Sci. Technol. B* **10**, pp. 2211-2216, 1992.
32. R. B. Marcus and T. T. Sheng, "The Oxidation of Shaped Silicon Surfaces," *J. Electrochem. Soc.* **129**, pp. 1278-1282, 1982.
33. D.-B. Kao, J. P. McVittie, W. D. Nix, and K. C. Saraswat, "Two-Dimensional Thermal Oxidation of Silicon - I. Experiments," *IEEE Trans. Electron. Devices* **34**, pp. 1008-1017, 1987.
34. Y. Uenishi, M. Tsugai, and M. Mehregany, "Micro-Opto-Mechanical Devices Fabricated by Anisotropic Etching of (110) Silicon," in *Proc. IEEE* Workshop on MicroElectroMech. Syst. (MEMS'94), Osio, Japan, Jan. 1994, pp. 319-324.
35. D. L. Kendall, W. P. Eaton, R. Manginell, and T. G. Digges, Jr., "Micromirror Arrays Using KOH:H_2O Micromachining of Silicon for Lens Templates, Geodesic Lenses, and Other Applications," *Opt. Eng.* **33**, pp. 3578-3588, 1994.
36. N. C. Tien, O. Solgaard, N.-H. Kiang, M.Daneman, K. Y. Lau, and R. S. Muller, "Surface-Micromachined Mirrors for Laser-Beam Positioning," in *Proc. Int. Conf. Sensors and Actuators* (Transducers'95), Stockholm, Sweden, June 1995, pp. 352-355.
37. D.-B. Kao, J. P. McVittie, W. D. Nix, and K. C. Saraswat, "Two-Dimensional Thermal Oxidation of Silicon - I. Experiments," *IEEE Trans. Electron. Devices* **34**, pp. 1008-1017, 1987.
38. J. J. Yang, M. Sergant, M. Jansen, S. S. Ou, L. Eaton, and W. W. Simmons, "Surface-Emitting GaAlAs/GaAs Linear Laser Arrays with Etched Mirrors," *Appl. Phys. Lett.* **49**, 1139, 1986.

39. J. J. Liang and J. M. Ballantyne, "Self-Aligned Dry-Etching Process For Waveguide Diode Ring Lasers," *J. Vac. Sci. Technol. B* **12**, 2929, 1994.
40. K. K. Ko, K. Kamath, O. Zia, E. Berg, S. W. Pang, and P. Bhattacharya, "Fabrication of Dry Etched Mirrors for $In_{0.2}Ga_{0.8}As$/GaAs Multiple Quantum Well Waveguides Using an Electron Cyclotron Resonance Source," *J. Vac. Sci. Technol. B* **13**, pp. 2709-2713, 1995.
41. I. Vurgaftman and J. Singh, "Steady State amd Transient Characteristics of Microcavity Surface-Emitter Lasers with Compressively Strained Quantum-Well Active Regions," *IEEE J. Quant. Electron.* **31**, pp. 852-863, 1995.
42. K. K. Ko, S. W. Pang and M. W. Cole, "Evaluation of Surface Damage on GaAs and GaInAs Etched with an Electron Cyclotron Resonance Source," *J. Vac. Sci. Technol. B* **12**, pp. 3382-3387, 1994.
43. M. Takahashi, Y. Michitsuji, M. Yoshimura, Y. Yamazoe, H. Nishizawa, and T. Sugimoto, "Narrow Spectral Linewidth 1.5 μm GaInAsP/InP Distributed Bragg Reflector (DBR) Lasers," *IEEE J. Quantum. Electron.* **25**, 1280, 1989.
44. J. Singh, Semiconductor Optoelectronics: Physics and Technology, McGraw-Hill, New York, 1995, pp. 537.
45. A. L. Gutierrez-Aitken, H. Yoon, and P. Bhattacharya, "High-Speed InP-Based Strained MQW Ridge Waveguide Laser," *Proc. Inter. Conf. on InP and Rel. Mat.* 476, 1995.

CHAPTER 4

Focused Ion Beams for Micromachining and Microchemistry

Diane K. Stewart
J. David Casey, Jr.
Micrion Corporation

CONTENTS

4.1 Basic Concepts of Focused Ion Beam Technology / 155
  4.1.1 Introduction / 155
  4.1.2 Liquid metal ion source / 155
  4.1.3 Typical column and systems / 156

4.2 Basic FIB Processes / 157
  4.2.1 Image acquisition / 157
  4.2.2 Milling / 158
    4.2.2.1 Description / 158
    4.2.2.2 Limits of milling / 159
    4.2.2.3 Gas-assisted etching / 160
  4.2.3 Deposition / 164
    4.2.3.1 Description of FIB deposition / 164
    4.2.3.2 Recipe for FIB deposition / 164
    4.2.3.3 Material composition / 166
    4.2.3.4 Limits of deposition / 166

4.3 Applications, Examples, and Techniques Unique to FIB Micromachining / 169
  4.3.1 Mask repair / 169
    4.3.1.1 X-ray masks / 172
  4.3.2 Circuit modifications / 173
    4.3.2.1 Examples / 173
    4.3.2.2 Endpoint detection milling technique / 175
  4.3.3 Failure analysis / 176
    4.3.3.1 Cross-sectioning / 177
    4.3.3.2 Decorative etch of a cross section using GAE / 178
    4.3.3.3 Exposure of buried metal conductors / 179

- 4.3.3.4 Vias / 180
- 4.3.4 TEM sample preparation / 181
  - 4.3.4.1 Techniques and examples / 181
  - 4.3.4.2 TEM recipe / 182
- 4.3.5 3-D microstructures / 186
  - 4.3.5.1 Example / 187

4.4 Summary / 188
- 4.4.1 Future / 188
  - 4.4.1.1 Nanomachining and FIB lithography / 189
  - 4.4.1.2 In-line process / 189
  - 4.4.1.3 Gas chemistries / 190
- 4.4.2 Limitations / 191
  - 4.4.2.1 Beam resolution / 191
  - 4.4.2.2 Gallium implantation / 191
- 4.4.3 Beyond gallium (liquid metal ion source) gas field source / 191

References / 193

4.1 BASIC CONCEPTS OF FOCUSED ION BEAM TECHNOLOGY

4.1.1 Introduction

Focused ion beam (FIB) tools were commercially introduced more than 10 years ago. The first tools were specifically designed to image features on lithographic masks and to repair defects to produce a defect-free mask[1]. The uses for the same basic tool have expanded to include modification of advanced integrated circuits (IC)[2]; failure analysis of high-performance devices[3]; transmission electron microscopy (TEM) sample preparation[4]; direct writing of three-dimensional structures on a microscopic and nanoscopic level, referred to as "microstructuring"; and formation of fine, three-dimensional probe tips for use in scanning probe microscopes[5]. In this chapter, we review the basic concepts of FIB technology and application, describe the complex microchemistries that enable advanced micromachining, and provide examples of mask repair, circuit modification, failure analysis, TEM sample preparation, and creation of microstructures. We conclude with a discussion of the future for FIB approaches in micromachining and the fundamental limitations of the technology.

4.1.2 Liquid Metal Ion Source

The heart of an ion beam tool is the source. For most, or almost all, commercial FIB tools, a liquid metal ion source (LMIS) that emits positively charged gallium ions accelerated in increments of 5 kV from 10 to 50 keV is used. The column typically consists of a gallium liquid metal ion source, electrostatic demagnifying lenses, deflectors, and blanking plates. The current originates in a liquid gallium ion source composed of a 1500+ hour gallium reservoir and etched needle assembly (Figure 4.1). User-selectable apertures in the column provide for a range of spot sizes and beam currents from 5 nm to 1000 nm, and 12 pA to 4 nA. Figure 4.2 is an FIB image of a sample of gold grains typically used for scanning electron microscope (SEM) image resolution; submicron features are easily resolved. The resolution of an FIB column is becoming equivalent to that of an SEM.

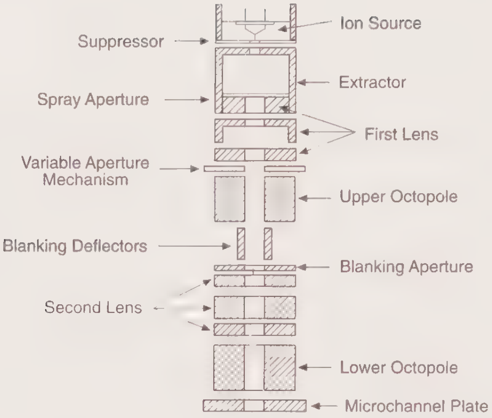

Suppressor
Ion Source
Extractor
Spray Aperture
First Lens
Variable Aperture Mechanism
Upper Octopole
Blanking Deflectors
Blanking Aperture
Second Lens
Lower Octopole
Microchannel Plate

FIG. 4.1. Typical LMIS FIB column optics diagram

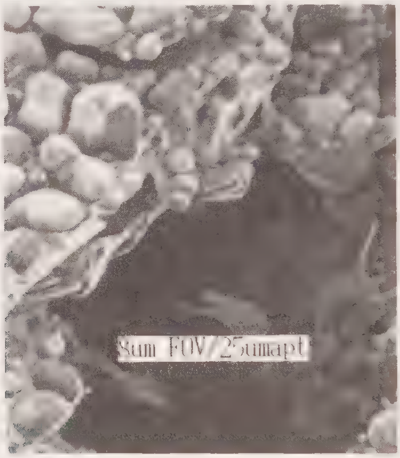

8um FOV/25umapt

FIG. 4.2. Secondary ion image created using FIB showing small features which can be resolved, a measure of how good the FIB is.

Thus the FIB can not only provide image information similar to that of an SEM but can also be used to construct microstructures.

The ion beam is digitally rastered across an area ranging from a few square micrometers to approximately a square millimeter to (1) acquire an image based on collection of secondary charged particles by either a microchannel plate or a channeltron electron multiplier, (2) to remove unwanted material via energetic or chemically enhanced sputtering, or (3) to deposit metal, carbon, or dielectriclike films in submicron patterns.

4.1.3 Typical Column and System

Figure 4.3 shows a typical ion beam system. The basic components include:

1. a focused ion beam column which is typically differentially pumped to $< 10^{-7}$ torr to protect the gallium source from contamination during chemical processes in the vacuum chamber,
2. a vacuum system with a typical base chamber pressure of $1'3 \times 10^{-7}$ torr which rises to $>6 \times 10^{-6}$ torr during chemical processes,
3. a substrate load lock to enable substrate handling,
4. operating system, consisting of a mouse and/or knob panel control of the system,
5. electronics rack, and
6. chemicals for microchemistry which are contained externally for those which require specific safety and pressure controls, and are metered to the chamber via feedthroughs into the vacuum system, or internally, which is preferred for chemicals which require heating.

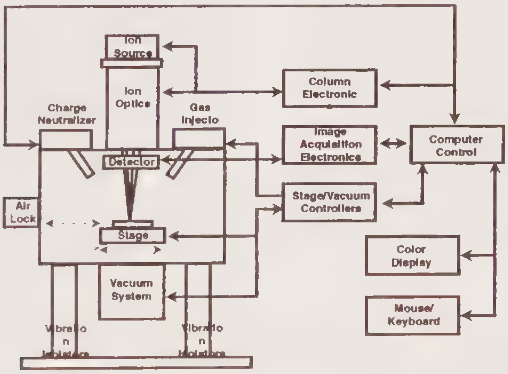

Ion Source
Ion Optics
Column Electronic
Charge Neutralizer
Gas Injecto
Image Acquisition Electronics
Computer Control
Detector
Air Lock
Stage/Vacuum Controllers
Stage
Color Display
Vacuum System
Mouse/ Keyboard
Vibratio n Isolators
Vibratio n Isolators

FIG. 4.3. Schematic of FIB.

Many FIB systems also have a multiaxis stage to provide tilt and rotation for precise sample positioning, high-resolution cross-sectional imaging, and device modification. Detailed descriptions of typical FIB systems can be found in the literature.[2,6]

4.2 BASIC FIB PROCESSES

4.2.1 Image Acquisition

An ion beam is used to generate a submicroscopic image of a substrate by digitally rastering the beam in increments across the surface to generate secondary charged particles that are collected on a charged particle detector. The number and type of secondary particles is a function of the material being sputtered, as well as the charged state of the sample surface. Positively (ions) or negatively (electrons) charged particles are collected on an appropriately biased detector such as a microchannel plate, channel electron multiplier, or Everhart Thornley (ET) detector to generate secondary ion microscope (SIM) or secondary electron images (SEM), respectively. Though the SEM images produced by FIB have a higher signal to noise ratio (S/N) than SIM images, SIM images provide much greater contrast due to the relative differences in secondary ions generated from different materials. Both types of images are useful as well as complementary. The acquired image can be stored for further analysis.

4.2.2 Milling

Milling a structure can be thought of as overimaging, or conversely, taking an image undermills a structure. However, to mill (or sputter) a discrete structure requires knowledge of the physics of sputtering and the parameters of the system.

4.2.2.1 Description

The rate at which material is milled or removed by the ion beam is proportional to the beam size and beam current. The amount of material removed is proportional to the dose delivered (unit charge/area). A rule of thumb is that a dose of 4 nCμm^2 is sufficient to remove 1 μm^3 of material. This is an adequate guideline for milling holes with aspect ratios of less than 5:1. Also, different materials have different sputter rates; the sputter yield of gold, for example, is more than ten times greater than that of tantalum. Compound materials such as indium phosphide mill nonuniformly, leaving phosphorus-rich nodules, while the indium-rich matrix is removed more easily.

What are the implications for micromachining? In general, each user of an FIB tool has a limited number of applications that are routinely required, although the number of applications is limitless. An FIB system can be configured with a set of files (parametric recipe cards) that take into consideration beam parameters and milling requirements for particular tasks. For example, lithographic mask makers routinely repair submicron defects on masks using a low-current beam (e.g., 25-nm, 25-pA diameter beam current) to mill or remove excess material from the layer of material that blocks the incident radiational. On an optical mask, the excess material is either chromium or a molybdenum silicide compound of uniform thickness of about 1000 Å across the surface, and on an X-ray mask, the absorbing layer is either gold or tantalum, 5000 Å thick. The defects (excess material) which are typically repaired by FIB are usually less than 2 μm^2. The only interaction required by the FIB operator in this example is to outline the structure that needs to be micromachined. Beam parameters and dose are coded into a job file. For example, Figure 4.4a is an image of an excess gold absorber on an X-ray mask; the defect has been outlined by FIB drawing tools, and the beam will be rastered over the area to remove the unwanted structure. Typical submicron defects are removed in a few minutes (Figure 4.4b).

In contrast, TEM sample preparation, requiring removal of many cubic micrometers of material, may take more than 1 hr of milling time with a high-current ion beam of 4 nA or greater. However, a simple recipe for TEM sample preparation can be stored in the operating system, and only the dimensions of the micromachined structure need be specified by the operator.

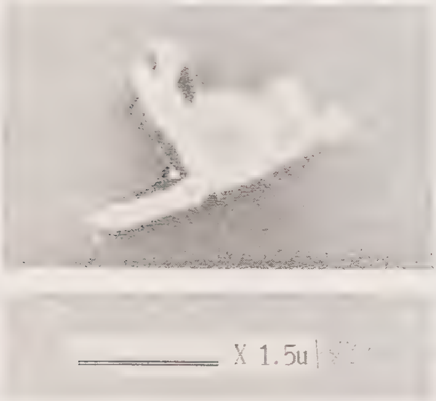


FIG. 4.4a. Image of defect on an X-ray mask outlined by polygon-shaped repair box

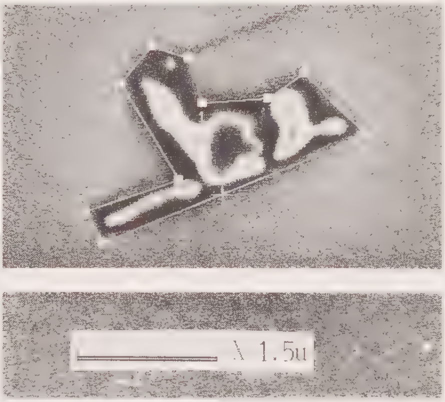


FIG. 4.4b. Image of defect from Fig. 4.4a, after dose of 2 nC/μm^2; bright areas indicate incomplete removal of material.

4.2.2.2 Limits of milling

The accuracy of beam placement and material removal is limited by the beam size as well as by the spatial density of features in the local area where structures have been machined. For example, accurate removal of excess gold on X-ray masks is constrained when features are closer together than 0.25 μm because of the redeposition of sputtered material that accumulates on adjacent structures.[7]

Milling deep, narrow holes with an ion beam to access buried metal circuitry is a typical example of where FIB milling reaches its limits rather rapidly. It is possible to sputter a 1:5 hole (e.g., 1 μm x 1 μm x 5 μm deep). Attempting to mill a greater aspect ratio hole than that with an FIB, no matter how small the beam diameter, creates a V-shaped base, and the angle of the side walls decreases to much less than 90 degrees.

4.2.2.3 Gas-assisted etching

One of the reasons that the aspect ratio of holes is limited is that the ion-sputtered debris redeposits faster on nearby features than the by-products can be pumped away into the vacuum system. Injecting a halogen-containing gas locally onto the area being sputtered can produce volatile products which are removed faster than they can be preferentially absorbed.[8] Gas-assisted etching (GAE) is a microchemical process; the chemical effects occur within the vicinity of the ion beam in microscopic dimensions to enhance micromachining. Using GAE with the FIB improves the aspect ratio of holes that can be milled, improves sidewall angles closer to 90 degrees, and enhances the removal rate of the process by removing more material for less beam time on the surface. In addition, GAE is material selective: different gas compounds etch different materials with different rates (Table 4.1).

TABLE 4.1 Typical etch enhancement factors[9]

Gas Material	Aluminum	Silicon	SiO_2	Tungsten
Br_2	8–16	5-6	0	0
Cl_2	7-10	0	0	0
XeF_2	0	7-12	7-10	7-10

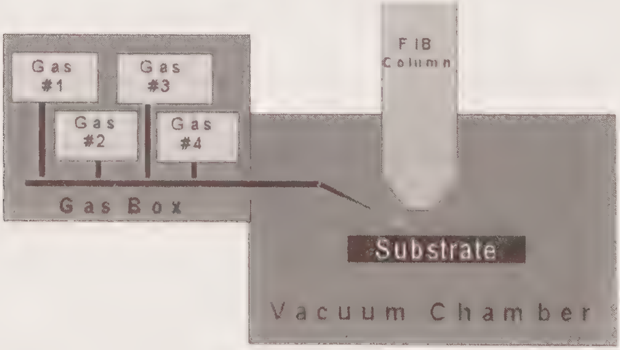


FIG. 4.5. Schematic of system configured with gas etching hardware.

The use of iodine has been described by Thaus et al. for enhancing removal of Ni-Fe alloys.[10]

Chemical reservoirs containing 10-20 g of precursor chemical may be either internal or external to the FIB chamber, but external delivery of chemicals allows for greater flexibility in cylinder interchange as well as safer handling of potential chemical hazards. Figure 4.5 is a schematic of typical GAE hardware configuration.

An example of GAE is shown in Figure 4.6a, where a 5 μm x 10 μm area was etched in aluminum without using gas enhancement. Etching with chlorine (Figure 4.6b) shows the same shaped hole with sharper walls and a smoother silicon dioxide floor.

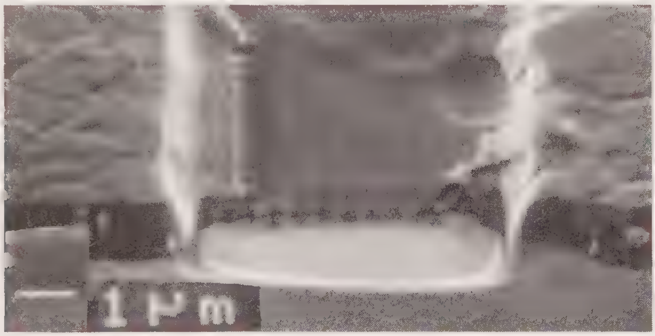


FIG. 4.6a. Image of an area of Al on SiO_2 removed by simple FIB sputtering.

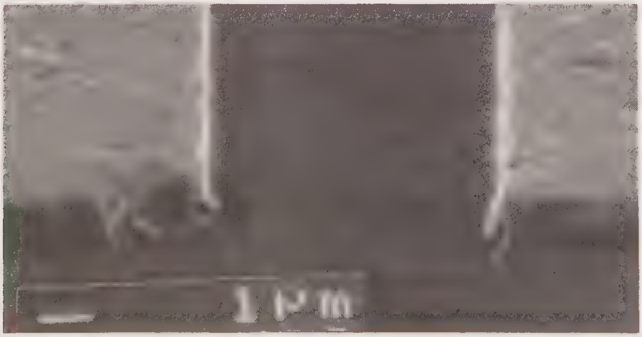


FIG. 4.6b. Image of an area on the same sample as in Figure 4.6a using Cl_2-GAE to remove Al. Note the sharper side walls and smoother SiO_2 floor.

In addition, a smaller ion dose was required with GAE so the overall process was faster: a smaller ion dose generally means less potential sample damage, as well as charging. Figures 4.7a, 4.7b, and 4.7c demonstrate the sequential application of GAE chemistries for exposing circuitry. In Figure 4.6a, XeF_2 was used to remove a 7-μm-thick layer of polyimide passivation to expose the aluminum lines. This is typically required to expose electrical circuitry under resist to allow for electrical probing, or simple device modifications such as milling breaks in electrical lines to open a line, or to deposit a metal-containing film to add a conductive path. Figure 4.7b shows a 25-μm x 20-μm window cut uniformly through aluminum to expose the underlying dielectric layer. Finally, the bottom metal layers were exposed by using XeF_2 to remove the interlevel dielectric (Figure 4.7c). The circuit remained operational for electrical probing.

FIG. 4.7a. A 40-μm x 45-μm hole opened using XeF_2-GAE through 7 μm of polyimide and passivation to expose the top metal circuitry.

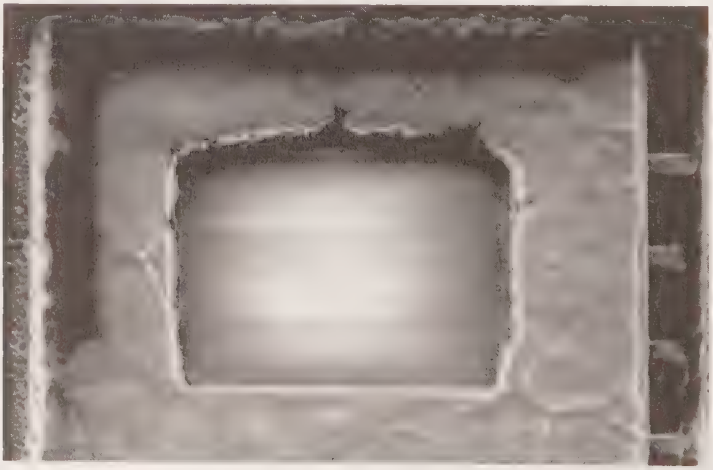


FIG. 4.7b. A 20-μm x 25-μm hole in metal opened using Cl_2-GAE to expose the underlying dielectric.

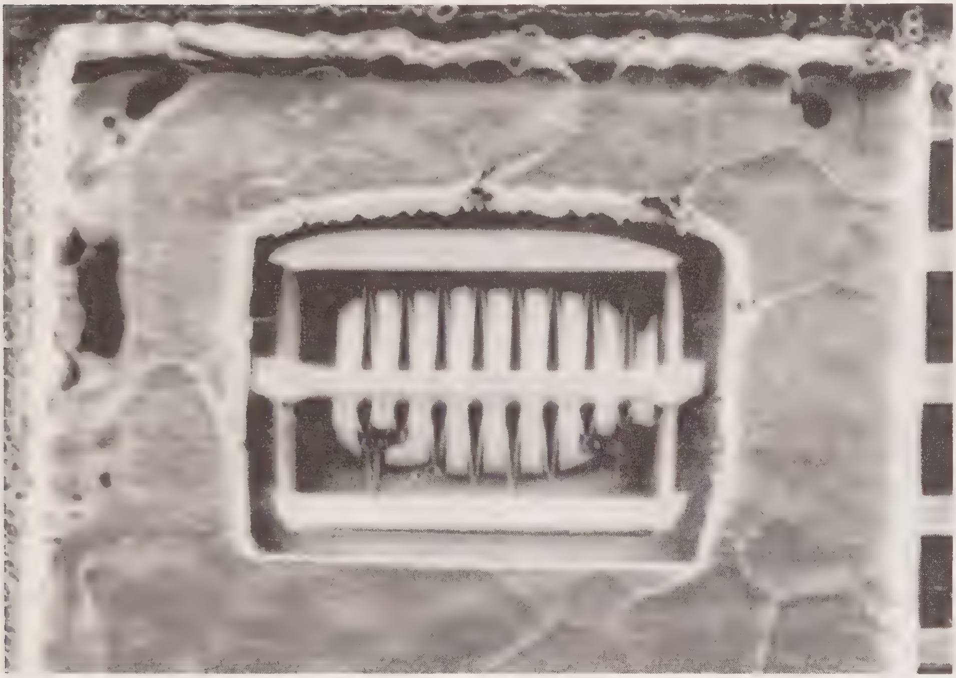


FIG. 4.7c. XeF_2-GAE was used to remove the dielectric and expose two underlying metal levels. The circuit remained operational for probing.

Most applications in device modification and failure analysis of integrated circuits (ICs) require selective removal of one material to expose another. Fig. 4.8 is an image of 1 μm x 10 μm holes created by using GAE-FIB to access a buried metal line. (The bright tungsten metal was deposited in the holes to outline the features.) Note the almost vertical side walls.

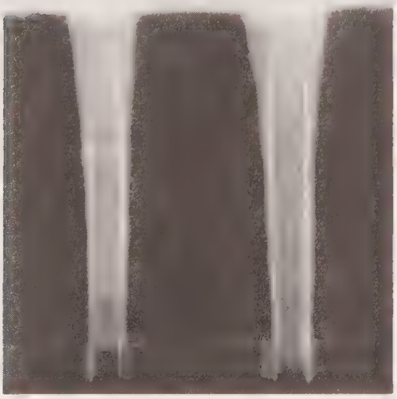


FIG. 4.8. High aspect ratio holes milled using GAE to go through metal and stop at the underlying dielectric layer. The holes are approximately 1 μm at the base. FIB-tungsten, which images bright, was deposited in the holes to delineate the side walls in the FIB image.

4.2.3 Deposition

4.2.3.1 Description of FIB deposition

In many applications, microscopic features can be added by microchemical vapor deposition. As in the GAE process, vapor from a reactive precursor is transferred to the surface and ejected via a small inner diameter (< 300 μm) needle positioned close to the sample surface. Typically the gas exit orifice is about 100-200 μm above the surface and about 300 μm from the center of the location of interest. The local gas pressure, and thus the concentration of gas available to adsorb on the surface, rapidly decreases as the distance between the orifice and the surface increases.

Mechanisms for FIB-induced material deposition from gas precursors have been described thoroughly in the literature.[11,12,13] For the purposes here, it is sufficient to know that net deposition occurs when more material is deposited than is sputtered. Thus, the goal in developing an efficient FIB deposition process is to enhance deposition at the expense of sputtering. This is best done by decreasing the effective beam current density at each pixel (digital beam element, or picture element). The controls to do this that are available on an FIB tool are beam current, dwell time (the time the ion beam spends at each pixel), and refresh time, which is the time the beam is intentionally kept away from the surface to allow adsorption of a monolayer of a gas. The dwell time should be kept as short as possible (< 1 μs), the refresh time is in the range of 2-5 ms, and the beam current density should be as small as is possible, generally < 5 A/cm.^{2} The point is to have as little beam on the surface as can easily be achieved using the system controls. Basically, a high current beam, dwelling a long time at each pixel, will mill the substrate faster than deposition can occur.

4.2.3.2 Recipe for FIB deposition

1. Bring the gas orifice to the surface as close as is reasonably possible without obscuring the features of interest in the image field. For example, the position of the orifice in x and y is about 300 μm from the center of the target, and the height above the surface is approximately 200 μm. The optimum position of the gas orifice may varying depending upon gas species and geometry of the repair (e.g. high aspect ratio hole filling requires a more normal incident of gas to enable more gas to diffuse into the hole, than deposition on a flat surface). For commercial systems, the manufacturer will have determined these parameters for their processes.

2. Choose a precursor chemical which:

a. Contains an atom of interest (for example, tungsten hexacarbonyl is used to deposit tungsten films).
b. Has a vapor pressure $\geq$/2 1 torr (the reservoir containing the precursor may have to be heated).
c. Can be handled safely (specifically chemicals which have hazard warning ratings greater than 3 out of 4 for health, reactivity, or flammability are not recommended).
d. Will "stick" (phyisorb) to the surface (i.e., have a sticking coefficient close to 1).

Unfortunately, item d is most difficult to predict. Harriott and Vasile[14] performed a complete study of the effect of side groups and "stickiness" of carbon-containing compounds for FIB deposition of carbon films for optical mask repair, but no similarly extensive studies have been performed for organometallic compounds.

Types of precursors

e. To deposit metals, many organometallics are worth trying, particularly the metal acetylacetonates, and the carbonyls (these can be liquids or solids). Metal films containing Au[15], Pt[16], W[17], Cr[18], and Al[19] have been deposited from such precursors.

f. Carbon films can be deposited from the polymer precursor families (such as the polystyrenes).

g. Films with dielectriclike properties can be deposited from siloxanes.[20,21,22] Oxygen can be preblended to increase the percentage of oxygen in the films.

Beam parameters

h. To deposit features less than 2 μm x 2 μm in area, a beam density of 5-10 A/cm^2 is required to avoid gas depletion (if the beam rasters too slowly over the surface, or the effective beam current density is high, the precursor gas does not have sufficient time to adsorb). The gas 'refresh' time is usually programmable; an increase in refresh time sometimes will improve the yield (number of atoms deposited per incident ion) of a deposition process as it allows more time for the gas to replenish. It is also essential that the beam dwell time at each discrete increment in the pattern be less than 1 μs for most deposition processes. Typically a submicron feature will take 1-3 min to deposit, depending on the yield of the process and the thickness desired. Gold deposits 40 times faster than tungsten, so thicker gold features can be deposited per unit time than can tungsten films.

More process latitude exists with larger features such as bond pads which can easily be written with beams as large as 0.5 μm in diameter with a 4 nA beam current without as much concern for gas depletion effects.

On a flat surface, the thickness of FIB-deposited material scales linearly with the ion dose. Using microchemistries to achieve high aspect ratio microstructures (e.g., GAE to open deep holes, tungsten deposition to fill holes) is not as predictable and requires a great deal of process development.

4.2.3.3 Material composition

Auger analysis of FIB-deposited metal films typically indicates that such films contain 10-20 at % implanted Ga and 10-20 at % C from the precursor or background residual contamination, the remainder being the material of interest. Film deposited under conditions to achieve high yield showed a greater percentage of included carbon.[15] Carbonlike films are composed of 70-80 at % carbon, 10 at % oxygen and 10-20 at % implanted gallium. FIB-deposited dielectric films are deposited from a combination of siloxane and oxygen, and show a 1:2 ratio of silicon to oxygen, and 10-20 at % Ga[22].

4.2.3.4 Limits of deposition

One of the biggest problems with performing accurate FIB deposition is the redeposition of sputtered material on nearby features. This is particularly serious when microstructures are deposited on lithographic masks to replace missing features. For example, a horizontal FIB-W film was deposited across an existing vertical line-space pattern in Figure 4.9. Excess sputtered material was redeposited on the sidewalls of the vertical (y direction) features so that the lines appear to 'bulge' in the direction of the FIB-W. This changed the critical dimension of the lithography. The broadened lines could be milled with the FIB to reshape the vertical lines.

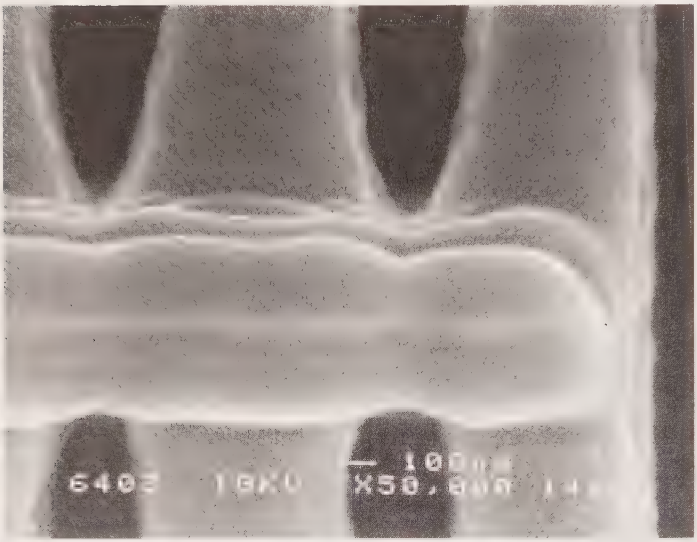


FIG. 4.9. SEM image of W film (tungsten) deposited across closely spaced, <1 μm-wide, vertical metal lines on a lithographic mask. Redeposition caused the local broadening of these lines, which is a problem in mask repair applications.

Improving the deposition yield (atoms deposited per incident Ga ion) improved the process because sputtering was reduced. However, the FIB tungsten deposition process has a yield of approximately 1 (at best), which is quite low and has a great deal of sputtering/redeposition inherent in the process. A higher yield (40-80) gold deposition process was developed by Wagner et al.[23], and in addition to producing the most dramatic high aspect ratio FIB features created, it provided a repair process with minimal redeposition for features spaced >0.25 μm (Figure 4.10). This has permitted the FIB mask repair approach to be extended to masks for deep submicron IC technologies.

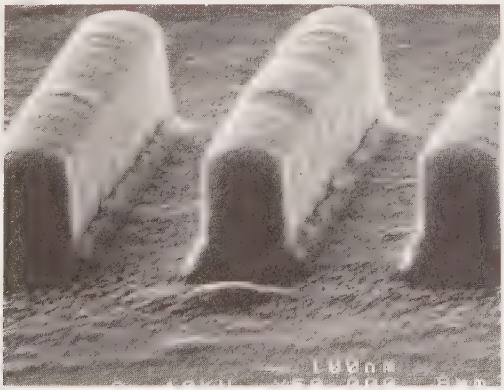


FIG. 4.10. SEM of 0.25-μm-wide gold lines deposited by FIB.

However, a higher yield process is not always the best for a given application. Some circuit repair applications require backfilling of high aspect ratio holes created by GAE with an insulator such as silicon oxide-based film. Typical insulator deposition processes have yields of 7-10[20]. Figure 4.11 shows FIB-milled areas which were not uniformly backfilled with an insulator. The voids were attributed to the "pent roof" (edge overhang limiting deposition at the base) phenomenon reported by Nakamura et al[20] in which faster growth at edges and corners leads to void formation.

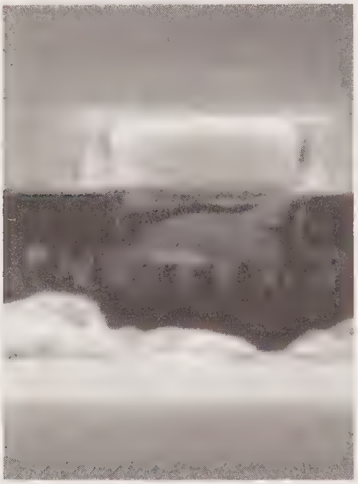


FIG. 4.11. FIB image of 2x2-μm^2 hole, 1 μm deep, milled to access metal line and backfilled with silicon dioxide material. The hole was cross-sectioned with the FIB to expose the voids.

Higher aspect ratio holes can be filled more effectively and with less voiding when the process yield is reduced. A typical recipe to reduce yield is to double the dwell time, effectively increasing the local beam current density. This allows holes of 5:1 to be filled. Figure 4.12 shows conformal coating of existing metal features with an FIB insulator deposited with a yield of <5 atoms/incident ion. Figure 4.13 shows a 5-μm-wide line of tungsten which was deposited to connect an aluminum line to an aluminum probe pad. There is a slight shadowing effect (material "piles up" on the left-hand side) of the features due to the position of the gas delivery needle. The low-yield tungsten (though a poor choice for mask repair) is well suited to filling high aspect ratio holes and to conformally deposit across features.

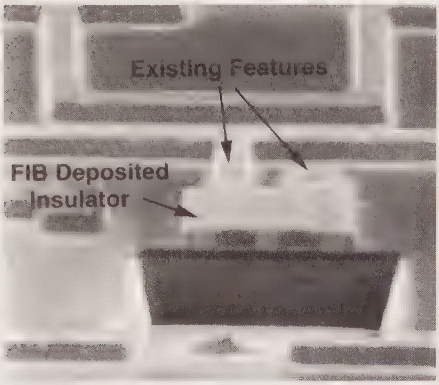


FIG. 4.12. FIB insulator deposited to uniformly coat existing metal features.

FIG. 4.13. FIB image of 5-μm-wide tungsten film deposited to connect an aluminum line to an aluminum pad for electrical probing.

4.3 APPLICATIONS, EXAMPLES, AND TECHNIQUES UNIQUE TO FIB MICROMACHINING

4.3.1 Mask Repair

Much of FIB technological development has been advanced by requirements for lithographic mask repair. A lithographic mask is the chip maker's stencil and can be reused thousands of times. Therefore repair of microscopic defects on a submicron-featured mask has economic merit. The repaired edge must not deviate from the intended position of the edge or feature by more than 0.10 μm in x and in y. The burden of the precision and repeatability, i.e., performance, is on the FIB system and beam. In this application, the FIB tool resembles an electron-beam writer more than a laboratory SEM.

Defects are removed by sputtering excess material or filled in by adding material. An optical mask (photomask) is typically a quartz plate with a pattern of thin (0.15 μm) chromium, which is optically opaque. The basic FIB recipe for chromium removal is a dose of 0.5 nC/μm^2 to remove 0.15 μm of chromium and a dose of ~0.15 μm-thick carbon film to replace missing chromium. (Refer to Figures 4.14a and 4.14b.) The edges of the repairs are within ± 0.1 μm of the existing edge.

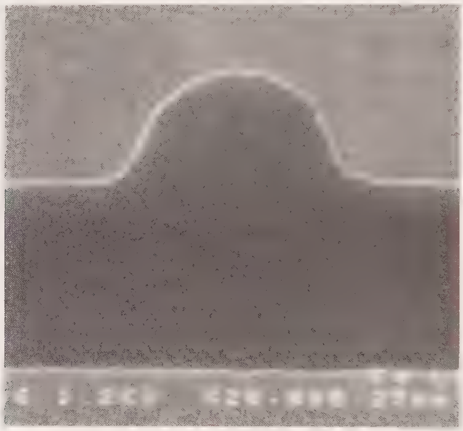


FIG. 4.14a. SEM image of missing defect on a chrome-quartz photomask before repair.

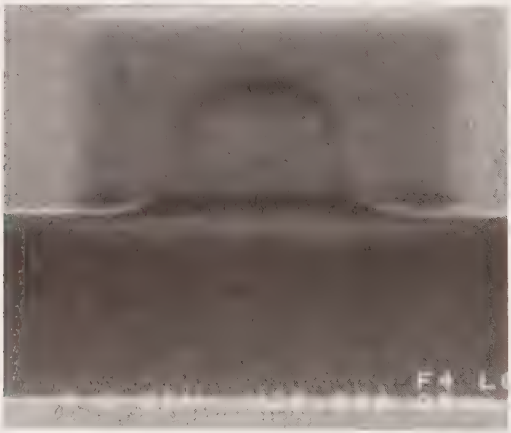


FIG. 4.14b. SEM image of defect in Fig. 4.14a repaired by depositing carbon to replace missing chrome.

The evolving generation of optical phase shift masks requires that micro-machined defect repairs be accurate in three dimensions. For a repair to be successful, it must: match size, shape, and position of the defect; reproduce the desired transmission; minimize damage to the underlying substrate; minimize damage to surrounding nondefect areas; and finally, the repair must be durable. Specifically, the defect must be repaired to within ±5% of the required thickness of the mask material. For example, a structure called a "phase bump," which produces a shift in light, can be selectively etched with a GAE process to fine tune the damage or removal of the quartz by the incident beam (Figures 4.15a and 4.15b). The degree of control in the z direction in FIB microchemistries has never been fully characterized and it is critical for repair of phase shift masks. More work needs to be done to characterize this.

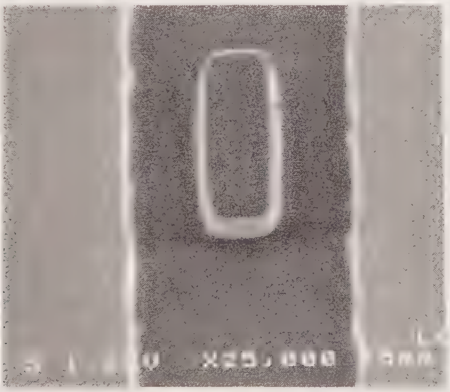


FIG. 4.15a. SEM image of a phase bump on quartz before repair.

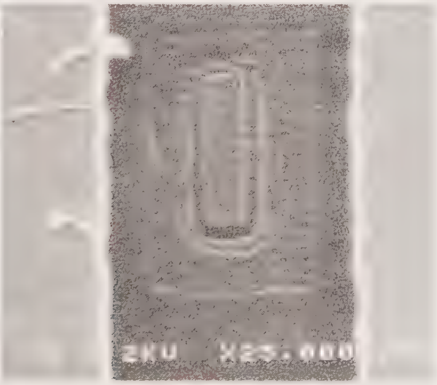


FIG. 4.15b. SEM image of the same area in Fig. 4.15a repaired by FIB milling followed by a low-dose etch of XeF_2 to remove a thin layer of gallium-implanted quartz. The gallium-implanted quartz reduces optical transmission, and removing it improves transmission.

4.3.1.1 X-ray masks

Although the Semiconductor Industry Association '94 roadmap[24] predicts that X-ray lithography will not be used in routine IC processing until after the year 2000, X-ray masks are considerably easier to repair than phase shift masks.

A simple X-ray mask is made by patterning a thick (>0.5 μm) metal X-ray absorber such as Au, Ta, or W on a thin (2-4 μm) X-ray-transparent membrane such as silicon or silicon nitride. Minimum feature sizes are <0.25 μm, implying removal or addition of high aspect ratio features. Gold, tantalum, and tungsten can be deposited to replace a missing absorber. Figure 4.16a is an FIB image of a silicon membrane patterned with 0.25 μm gold features. In the center of the figure is a defect where gold is missing. Figure 4.16b is an image of the same area after repair, which was performed by depositing gold from a gold organometallic precursor with a short dwell time, a long refresh time, and a 25-nm, 25-pA beam. The edges of the repaired defect were further delineated by trimming with the FIB to remove rdeposited material.

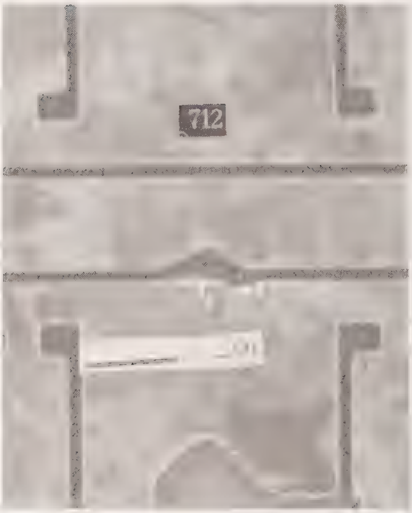


FIG. 4.16a. FIB image of a gold defect defect on an X-ray mask before FIB repair.

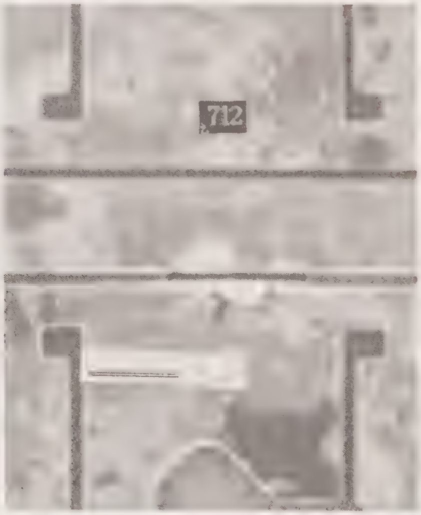


FIG. 4.16b. FIB image of the from Figure 4.16a after repair by gold deposition.

Unlike phase shift masks, X-ray masks are conductive and not prone to charging in the ion beam, which can affect repair placement. Further, the tolerances on the deposition thickness (7500 Å) and depth control of material removal, where the ion beam can remove 500-1000 Å of membrane, are easily achievable. The micromachining challenge is to deposit material precisely and with a high aspect ratio. Figure 4.17 shows FIB-deposited tantalum with the steep side walls required for X-ray mask repair.

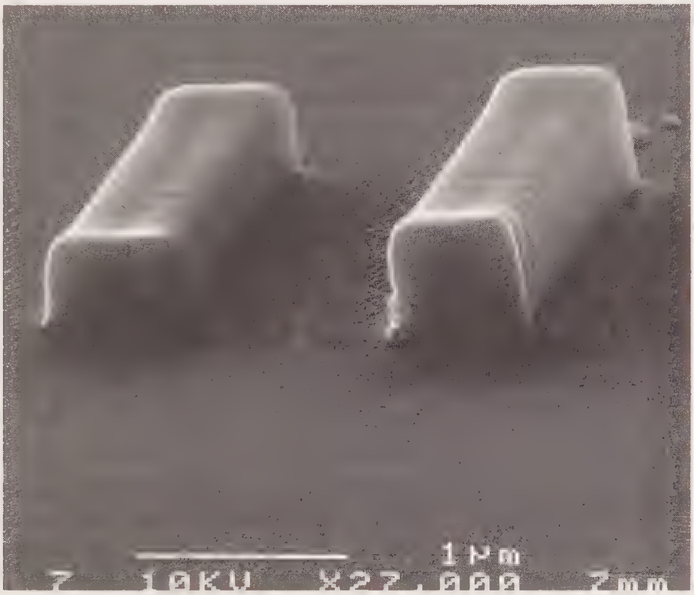


FIG. 4.17. SEM image of FIB-deposited tantalum films.

4.3.2 Circuit modification

4.3.2.1 Examples

FIBs are ideally suited to microrestructuring of advanced electronic devices. The top-level metal of devices can be locally rewired by micromachining precise submicron cuts or disconnects, or by adding low resistance connections by using FIB-deposited metals such as tungsten. The resistivity of FIB W is ~150 μΩ•cm, which is low enough to form conductive connections such as the interconnection shown in Figure 4.18. Figure 4.18 shows an example of using the FIB system to locally rewire the metal interconnects on an IC. In this example, the FIB is used to make a connection between metal 1 (M1) and metal 2 (M2), and to make isolation cuts to complete the rewiring. An ion dose of 15 nC/μm^2 was used to deposit 8 μm^2 of W interconnection shown in Figure 4.18b. The connection is precise and the contact holes were filled uniformly. In this case, the modified IC could be tested using an IC tester or parameter analyzer.

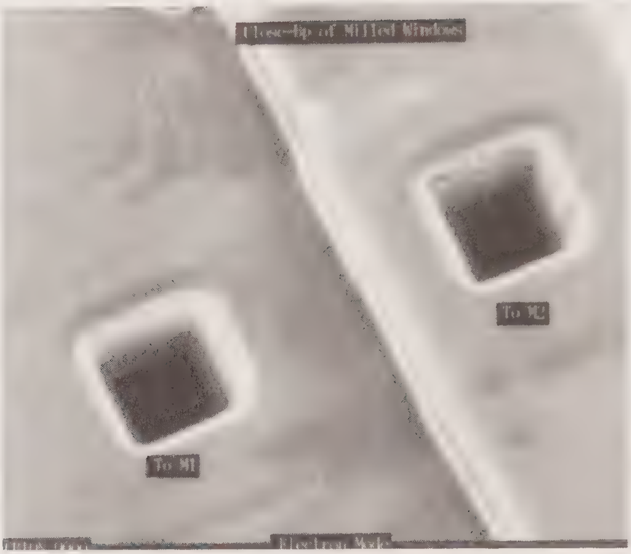

Close-Up of Milled Windows
To M2
To M1
Electron Mode

FIG. 4.18a. 2x2-µm windows milled through passivation to access buried metal.

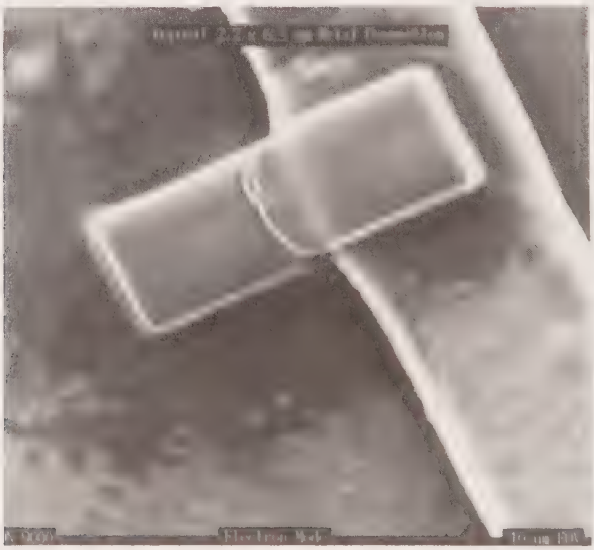

Electron Mode

FIG. 4.18b. Tungsten deposited to make connection shown in Fig. 17a. This figure also shows the isolation cuts.

Another example of using the FIB for device modification is shown in Figure 4.19. The passivated lines were isolated by milling seven separate cuts. Three lines were subsequently reconnected with FIB-deposited W; the resistance was measured and was within 10% of the calculated value for the new connection.

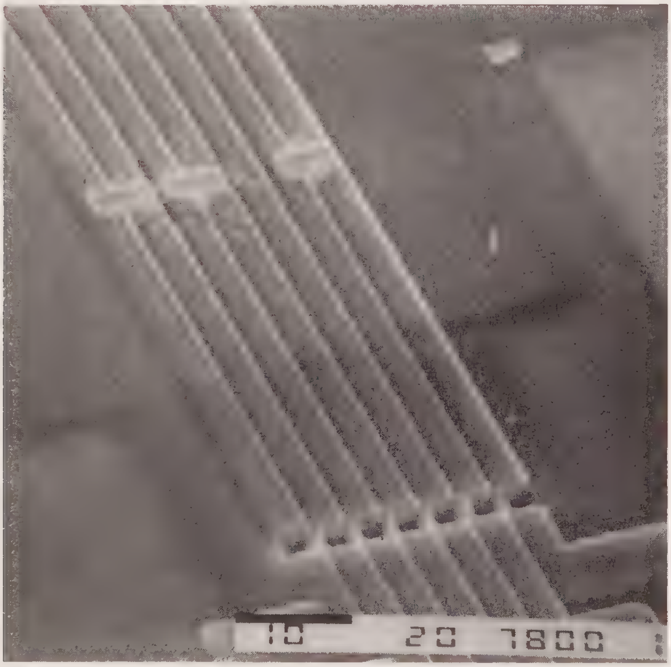


FIG. 4.19. Example of FIB circuit modification.

4.3.2.2 Endpoint detection milling technique

Endpoint detection and quantification of secondary ions emitted from materials during milling can be used to identify transitions between layers. The detection accuracy is better than 20 nm in the Ƶ direction on ICs and wafers. If an endpoint detection technique were not used, a contact hole milled too deeply could sever other lines, or if too shallow, would not expose the buried layer of interest.

Figure 4.20 shows the plot of secondary ion intensity vs. dose per area, with the device layer structure on the right. If the device layer material and thickness are known ahead of time, the plot information is sufficient to detail and record transitions through different materials. The intensity of the signal (which is plotted on the video screen during device modification) changes in real time as the beam mills through different materials. In addition to monitoring intensity of secondary ions or electrons as an endpoint detection technique, monitoring stage current is an accepted technique, as is secondary ion mass spectroscopy.

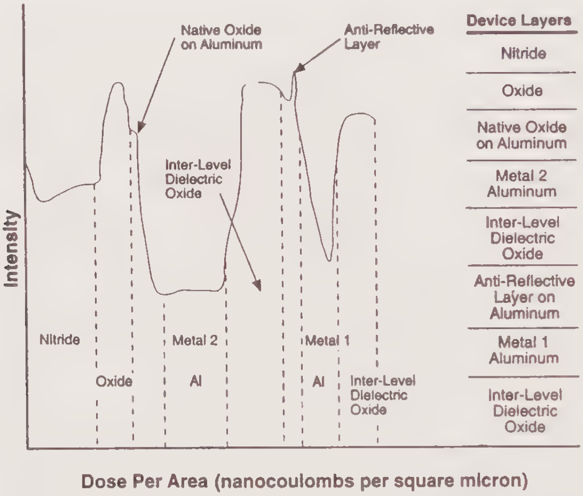


FIG. 4.20. Sample endpoint plot for milling through layers of different material; the intensity of the secondary ions generated from the sample changes as the beam goes through different materials.

4.3.3 Failure Analysis

Ideally, micromodifications should not alter nearby devices compared to existing failure analysis (FA) methods for sample preparation and defect isolation. Even though the material removal rate is relatively slow (a few cubic microns per minute maximum) the FIB approach for cross sections is much faster than the standard grind/polish procedure, and in some cases the device remains functional even after a cross-section. For example, samples cross-sectioned for SEM analysis require hours of iterative cleave-lap-decorative etch, conductive coat steps[3]. In contrast, a cross section 8 μm x10 μm by 6 μm deep can be prepared using an FIB in less than 30 min, and can be imaged in the same tool. Though material removal by lasers is faster than an FIB, the result is less precise. The spot size of a laser is at best 1 μm which is too large for many applications.

In this section we discuss several typical FA applications. Exhaustive case histories of FIB applications for FA are included in proceedings of the International Application Symposium for Testing and Failure Analysis (ISTFA) meetings and in proceedings of the International Reliability Physics Symposium (IRPS).

As IC geometries continue to decrease into the deep submicron range, FIB systems have become and will continue to be indispensable FA tools. Typical FIB approaches for FA include cross-sectioning of defects and structures, isolation of portions of a circuit for electrical probing, localized removal of passivation and/or deposition of probe pads for electrical or electron beam probing, and circuit modification (discussed in Section 4.3.2).

4.3.3.1 Cross-sectioning

A typical example of an FIB cross-section is shown in Figure 4.21. This is an image of a three-dimensional exposure of two faces cross-sectioned at 90 degrees.

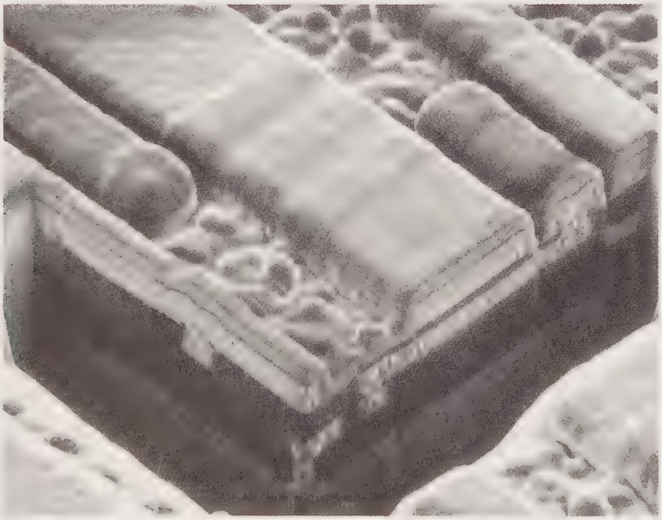


FIG. 4.21. FIB image showing 90-degree cross-section. Two 20-μm-long perpendicular cross sections were milled on the IC to provide a quasi, three-dimensional view of the device structure.

Recipe

1. Load the sample into the FIB chamber.

2. Navigate to the desired site using computer-assisted design (CAD) or other data base (data transfer of circuit layout is essential for location of features buried under passivation which can be difficult to image with the FIB). This is often not

available when looking at parts made by another manufacturer. Seeker holes, or other approaches, are needed in general.

3. A high-current beam (>300 pA in current, 0.15 μm in diameter) is used to mill two rectangles to intersect at the desired location. (The two large mills in Figure 4.20 took about 20 minutes each.)

4. A low-current ion beam (25 pA in current, 25 nm in diameter) is then used to remove thin slices from the two cross-sections to prepare a "polished" face for imaging. Each "polish" mill takes a few minutes.

5. The sample is then tilted (~45 degrees) and imaged in the FIB.

Although commercial SEMs have smaller spot sizes than FIBs and can therefore theoretically provide higher resolution images than an FIB, the FIB has an advantage in that different materials image with different contrast levels. (Metals look darker in secondary ion mode, which can be used to complement information obtained in secondary electron mode.) Each material has its own characteristic secondary ion yield generated by ion impact; thus it is easier to distinguish bright (metallic, or high-yield) layers from dark (dielectric, low-ion yield) layers on a cross section in secondary electron mode.

4.3.3.2 Decorative etch of a cross section using GAE

Determination of the thickness, uniformity, and quality of different materials in a multlayered device is an important IC analysis activity. The material contrast can be enhanced by selective chemical etch techniques which remove different materials at different rates. The sharper delineation of the layers provides higher resolution of the layers than can be achieved by SEM imaging alone. However, mechanical cross-sectioning combined with chemical etch techniques suffers from lack of repeatability, difficulty in precisely locating the feature of interest, and the fact that the technique is destructive to a large portion or all of the sample.

In contrast, a GAE process can be used in conjunction with the FIB to prepare and decorate a cross section more repeatably and quickly without destruction of the entire sample. Figure 4.22 is an FIB cross section of an IC decorated with a very low dose of XeF_2. The XeF_2 significantly enhances etching and thus recession of silicon and silicon dioxide layers, while leaving the metal lines less recessed. Lines as narrow as 7 nm have been delineated using GAE decoration.

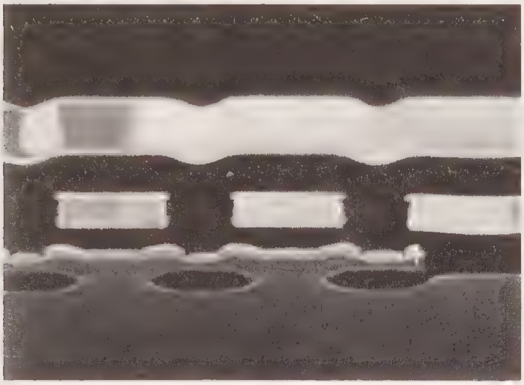


FIG. 4.22. FIB image of cross section of a circuit decorated with a low dose of XeF_2

4.3.3.3 Exposure of buried metal conductors

Complex electrical characterization of ICs is now possible by using a combination of FIB microchemistry and micromachining techniques to access buried metal conductors. In Figure 4.23, passivation was removed using XeF_2 FIB to expose M1 (the buried metal layer) and M2 (the metal layer above M1). From an FA point of view, this could be done to perform electron beam probing to determine if the device is functioning properly, or to look for structural defects.

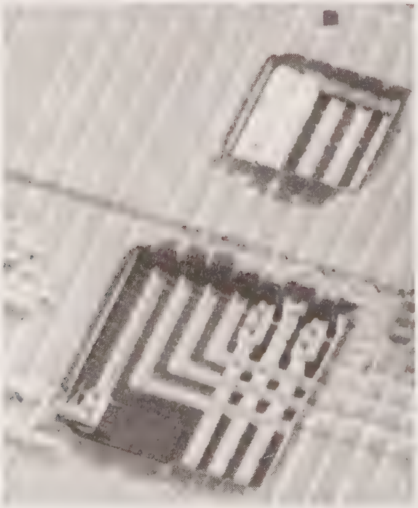


FIG. 4.23. FIB image of metal layers exposed by using GAE to remove passivation.

4.3.3.4 Vias

Milling high aspect ratio holes or vias to access buried metal is one of the more challenging applications of an FIB. The most complicated but frequently used procedure is:

1. A via is milled (typically 2 x 2 µm, several microns deep) to a buried metal line.

2. The via is backfilled with an FIB dielectric to passivate exposed metal lines.

3. A second opening (typically 1 x 1 µm) is milled in the middle of the dielectric.

4. FIB metal is deposited to connect the buried metal to a higher metal layer or to the surface for probing.

This procedure requires precision milling and precision and conformal filling. Conformal (or uniform) deposition of FIB dielectric is a challenge because the chemical precursors used to deposit insulators have a high yield and deposit very quickly on corners at the top of the high aspect ratio holes, causing voids in the middle of the deposits.[20,21] A technique has been developed to deposit a dielectric film into a hole without voids, in which the yield of the process is decreased during hole filling. This "recipe" is also applicable to conformally coated exposed geometries with high yield material.

Recipe for high aspect ratio hole filling and conformal line coating

1. Increase the dwell time per pixel (such as from 0.1 µs to >1.0 µs) to decrease the yield. A reasonable rule of thumb is that a 1x1 µm hole, 1 µm deep can be filled using a dose of 1 nC/µm^2.

2. Position a box at the center of the surface area to fill the hole; the deposit area should be 20-30% smaller than the hole to be filled to avoid edge effects.

3. Cap the filled hole with a higher yield process; the dwell can be decreased to <0.7 μs.

4. Exposed lines can be conformally coated by positioning dielectric deposition between lines and capping the structure with a coat of dielectric.[21]

Figure 4.24 shows a via made to connect a buried aluminum plug to the top surface for electrical probing. The darker silicon dioxide film adequately covered the side walls of the hole. The bright material in the center of the structure is FIB-W; although a void was left in the W deposition process, electrical continuity was confirmed for this connection.

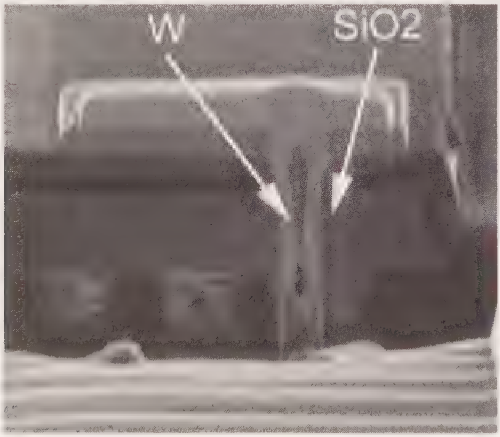


FIG. 4.24. FIB image of a multistep FIB connection made to buried metal.

4.3.4 TEM Sample Preparation

4.3.4.1 Techniques and examples

FIB micromachining has drastically reduced the time required to thin a sample for TEM analysis. The use of FIB provides the high-resolution imaging and precision thinning required to analyze specific features of complex ICs by TEM. The standard TEM sample preparation technique for cross sections of IC structures includes mechanical grinding and polishing, thinning (dimpling), and broadbeam ion milling. This approach lacks the precision (it is very difficult to produce a cross section of a particular feature of interest) and speed provided by FIB tools.

A typical TEM sample (mounted to cover the grid in the TEM) should be about 3000 μm long and 600 μm wide to cover the TEM grid, yet only 200-5000 Å thick. The sample must be thinned and then freed from the "parent" specimen. An excellent hands-on description of TEM sample mounting, FIB preparation, and subsequent manipulation is given by Lange and Czapski[25]. In this section we present a series of images (Figures 4.25a-d) taken during sample preparation that illustrate the different steps required.

4.3.4.2 TEM recipe

The system used in this TEM sample preparation study was a Micrion 9100EX. A gallium ion beam column provides for a range of beam current and spot sizes from 10 pA to 4nA and from 50 nm to 1000 nm. An add-on electron column was used to image the samples with a <50-nm beam. This two-column system provides for the ability to do high-precision milling and high-resolution imaging in one chamber.

1. Thin sample mechanically.

2. Mount the specimen normal to the ion beam.

3. Navigate to the site of interest (for example, a circuit failure) using CAD database information.

4. Position the area of interest in the center of the field of view. The sample can be precoated with FIB-W or Pt to prevent top surface damage if the area of interest is near the top surface.

5. Using a high-current ion beam, mill two coarse openings to define the region to be thinned (Figure 4.25a.) It may take approximately 2 hr to remove material using a beam with current greater than 300 pA depending on the size of boxes, the mill depth, the thickness of the "slab", etc.

6. Using a lower current beam (<300 pA) thin the sample from each side until the sample is <0.5 μm (5000 Å) (Figure 4.25b). Lower current beams provide precision thinning in addition to minimizing top surface damage from the "tails" (or low-beam current spread) of the ion beam.

   Thinning or polishing with a smaller ion beam may take as long as an hour (Figure 4.25b.)

7. The thinned TEM sample can be tilted in the 9100EX and imaged with the electron beam (Figure 4.25c.) Sample thickness is readily determined by inspection and the sample can be thinned again before removal from the vacuum chamber (Figure 4.25d.) TEM cross sections as narrow as 10 µm have been prepared so that several samples could be mounted simultaneously in a TEM.

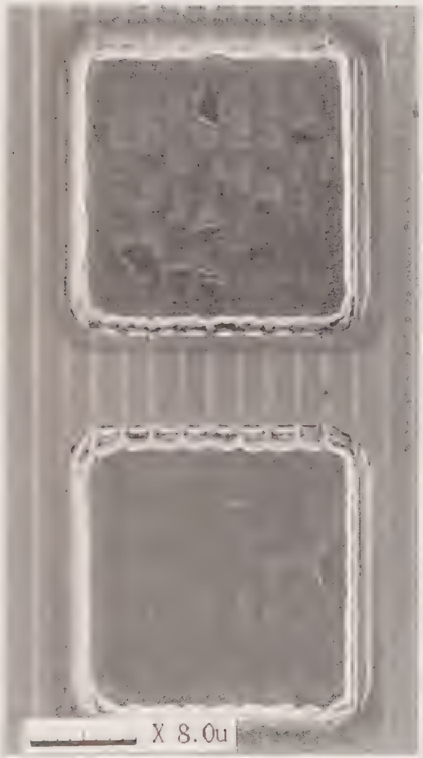


FIG. 4.25a. First step in TEM sample preparation by FIB. (Figs. 4.25 courtesy Kate Noll, Micrion Corp.)

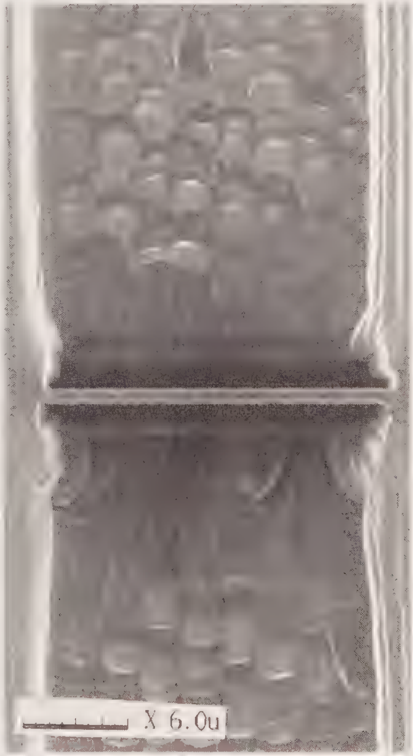


FIG. 4.25b. Thin section of TEM sample after coarse milling and prior to final "thinning" using a small-diameter, low-current ion beam.

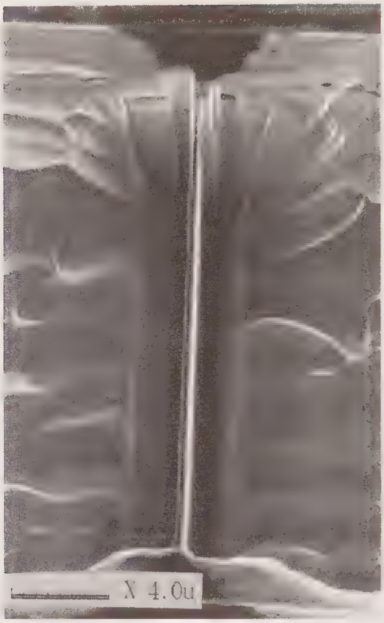


FIG. 4.25c. Tilted image of sample prepared for TEM analysis.

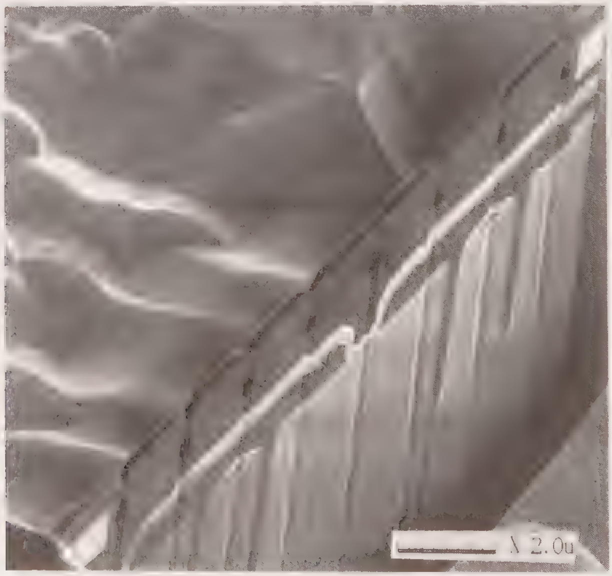


FIG. 4.25d. Top view of TEM sample thinned to <0.5 μm using FIB.

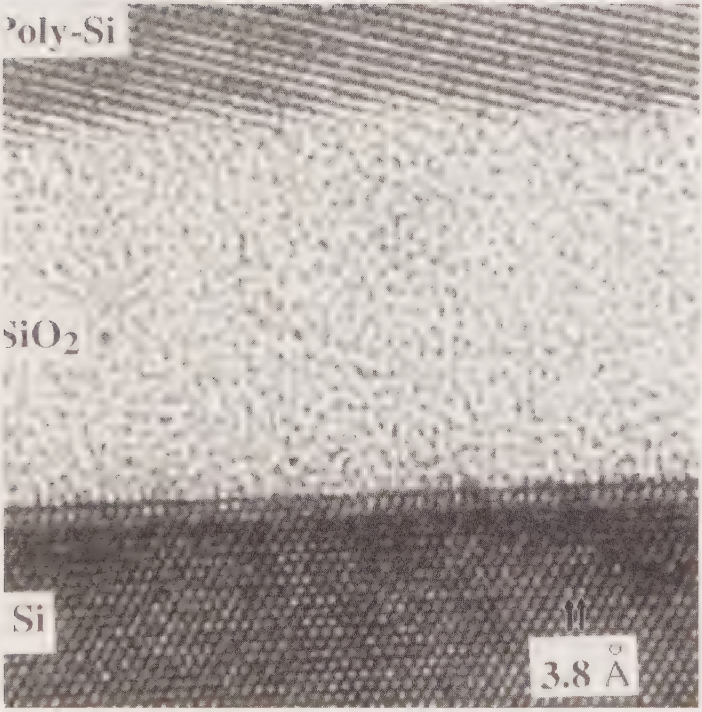


FIG. 4.26. A high resolution TEM image showing the crystalline structure of the Si substrate and polysilicon as well as the amorphous structure of the SiO_2.

Figure 4.26 is a TEM micrograph of a layered microelectronic structure; the sample was prepared by FIB. The crystalline lattice is intact and shows that FIB is a viable technique for TEM sample preparation.

One way of reducing the milling overhead time on a TEM sample is to use GAE during the coarse etching. Chlorine GAE has been used to thin gallium arsenide (Figure 4.27) and indium phosphide. These materials have very nonuniform sputter rates due to the difference in sputter yields for gallium, indium, and phosphorus. Frequently, phosphorus nodules protrude after sputtering. However, chlorine GAE enhanced mill rates and provided a much smoother surface finish than the FIB alone on these particular materials.

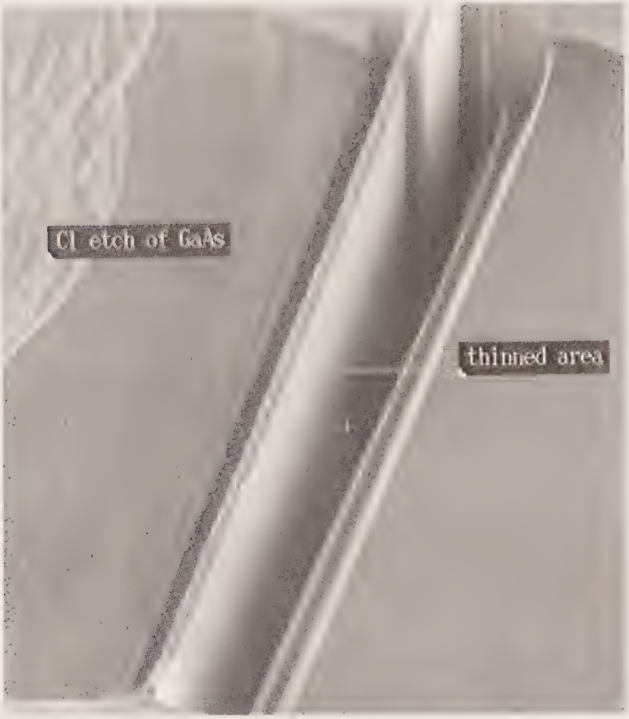


FIG 4.27. FIB image of GA sample prepared for TEM by Cl_2 GAE.

4.3.5 3-D Microstructures

Prototyping three-dimensional microstructures is possible with FIB milling. Micromechanical structures are now routinely produced in quantity by using X-ray lithography techniques and standard silicon processing[26]. FIB is not in competition with this technology nor should it be expected to compete. The limitations are the time required to micromachine a single structure and the inability to create more than one structure at a time with a single beam (3-D writing is slow and serial, and is not practical for creating multiple structures sequentially).

Why take the FIB approach? First, features can be customized and prototyped before tooling up to create multiples using more standard techniques, much as FIBs are used by IC design engineers to modify and test ICs. Second, asymmetric features can be machined such as the unique hooks created by Vasile and reported in Ref. 5. Many of the tips described in this article were for novel scanning probe microscopy (SPM) applications. The technique is not useful for batch production of tips, but in this case fabricating probe tips using FIB is a practical approach since the dimensional requirements are more important than the actual number of tips produced. One of the other performance requirements an FIB system must meet is the ability to deflect the beam in a circular pattern. Although it is possible to rotate the sample during preparation much as a lathe would do, deflecting the beam allows for milling more precise and smaller features.

4.3.5.1 Example

An example of a micromachined probe tip is shown in Figure 4.28a. A coarse tip on the silicon nitride cantilever arm was obtained by a different process. The goal was to make the dimensions of the tip as small as possible using an FIB. The milling pattern was annular. A high-current beam (300 pA, 0.1 μm) was used to mill a square, leaving an open square in the middle to define the tip. This took 3 min. Then a smaller beam, 25 nm, was deflected in an annular circle to sputter the central rectangular probe into a conical probe. This also took only a few minutes. Doses were less than 1 nC/μm^2 due to the high sputter yield of the material, as well as the shape. The finished structure (Figure 4.28b) is 1 μm at the base and approximately 8 μm tall. The jagged peaks at the base result from nonuniform milling and redeposition.

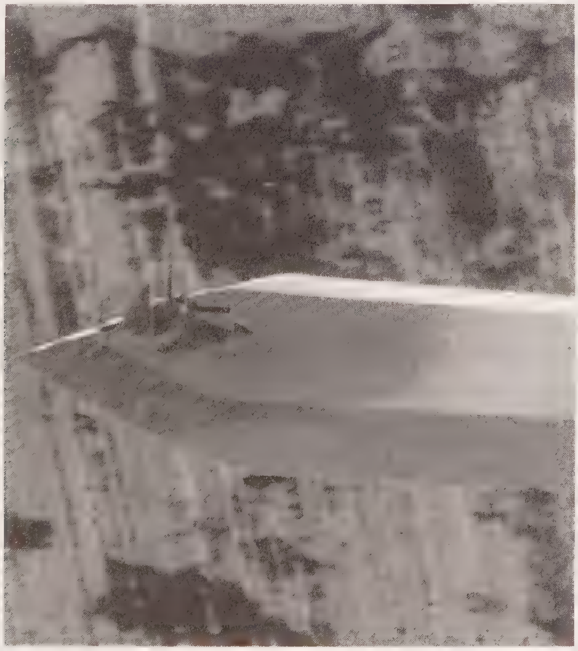


FIG. 4.28a. FIB image of micromachined probe tip on end of silicon nitride cantilever. (Courtesy R. Lee, Micrion Corp.)

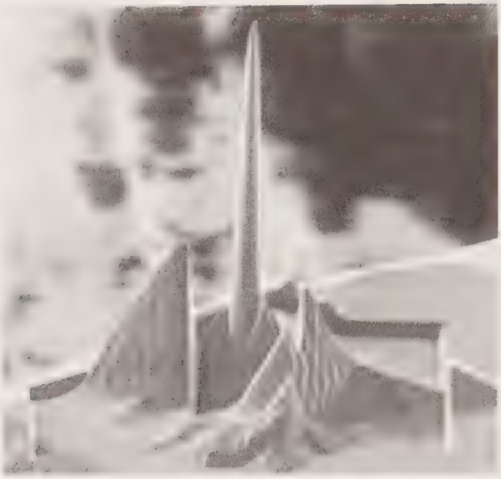


FIG. 4.28b. Closeup of probe tip from Figure 4.28a. The base of the tip is about 1 μm and the tip is 8 μm.

Researchers have attempted to fabricate probe tips by using FIB-deposited conductors. The 1x15-μm gold pillars deposited by Blauner[7] et al. represent the highest aspect ratio FIB depositions ever published. However, the high carbon content (>40%) of most FIB metals make them unlikely candidates for probe tips.

4.4 SUMMARY

4.4.1 Future

In 10 years, FIBs have achieved a status in microelectronics research, development, and manufacturing support laboratories similar to that of an SEM tool. They are the backbone of prototype micromachining and modification of electronic devices, repairing lithographic stencils, and creating novel microstructures. A number of microchemical processes have been developed to enhance milling rates and to deposit metals, carbon films, and dielectric-like films. One question is: Can the technology be advanced to provide nanomachining applications?

4.4.1.1 Nanomachining and FIB lithography

It is believed that an LMIS FIB column could in theory be made to produce a beam as small as 2 mm in diameter.[27] A reduction in ion beam diameter is desirable because the minimum feature size that can be imaged is directly related to the ion beam diameter. One application requiring FIB nanomachining is construction of nanoscale (i.e., machining using a nanometer scale ion beam) features typically fabricated using electron-beam lithography.[28] However, direct electron beam (EB) lithography is limited by electron scattering in the resist, which causes a spatial resolution.

FIB direct writing may have a limited role in nanolithography, but the limit is the writing speed. The FIB approach has been proposed but the writing speed is not competitive[28] with EB lithography. However nanoscale FIB patterns are useful for prototyping unusual nanostructures. An example of these are $GaAs/Ga_{1-x}Al_xAs$ heterojunctions in nanoscale dimensions formed by ion-beam implantation, which modulate the local electron energy profile.[29]

4.4.1.2 In-line process

The majority of the applications described in this chapter are applicable for R&D purposes, but generally are not easily automated for more routine applications. However, semiconductor manufacturers are beginning to appreciate the nondestructive precision milling and imaging capabilities the tool can provide for monitoring the steps of a wafer fabrication cycle. Typically, most wafer inspections are done off-line on a sacrificial wafer to monitor parameters such as the sheet resistivity of a layer or oxide thickness.

Cross-sectioning with the FIB provides a direct visual image of all layers stored in a single image frame of information. The ability of the FIB to provide material contrast information as described in this section is extremely valuable for monitoring film thickness, even layers as thin as 5'7 nm.

Typical In-Line Process Sequence[6]

Recipe

1. Remove a wafer from the fabrication line.
2. Insert into FIB.

3. Navigate to a predetermined test site. The stage in an FIB is capable of being driven to within <±0.5 μm of the desired site.

4. Expose the area of interest by cross-sectioning with the FIB; the size of a typical cross section required to provide adequate image information is about 4x4 μm^2 by 1-2 μm deep. A beam with current greater than 70 pA is adequate to perform the cross-section quickly (in several minutes).

5. A beam <25 nm in diameter is used to "polish" the face of the exposed circuit.

6. The sample is then tilted and a high-resolution image acquired and recorded or stored.

7. It is possible that the wafer could be returned to the fabrication line.

8. There is still an unresolved concern regarding the concentration of metal ions deposited per unit volume and the effects of implanted gallium on sites adjacent to the cross section. However, it is believed that the trace contaminants should not be appreciable.

4.4.1.3 Gas chemistries

It is likely that the resolution limits of FIB microchemistry have been reached, that is, most users do not attempt to deposit lines narrower than 0.2 μm. Further, a number of chemical processes have been developed that are adequate for most FIB applications. The degree of z resolution, or control and repeatability in deposit thickness or GAE etch depth, has not been established for many of the gas delivery systems in use. This is important for minimizing glass or quartz damage associated with repair of phase shift masks, where glass damage itself becomes a phase-shifting defect. More work needs to be done in this area. Also, the semiconductor industry has become more aware of the environmental impacts of the chemicals used in processing. Most of the compounds used in FIB chemistries, although consumed in quantities of <10 g/per year and used following OSHA guidelines, pose hazards to the environment. Therefore, it is likely that effort will be spent in identifying less hazardous precursors. For example, Wagner et al.[30] and Stark et al.[31] have shown that water vapor can be used with FIB to enhance the etch rate of photoresists and polyimide materials.

4.4.2 Limitations

There are fundamental limits to LMIS technology as well as limits imposed by gallium implantation.

4.4.2.1 Beam resolution

The fundamental limits that affect the resolution for micromachining applications are the ultimate size of a beam as well as the brightness, or beam current, that governs the writing speed. In theory, ion beams could be focused to smaller dimensions than the beam resolution[27] achieved by commercial SEMs. Technical issues such as the requirement to machine high-tolerance, high-precision electrostatic focusing elements in a column, will practically limit useful LMIS beams to spot sizes of 2-4 nm. Mechanical and electrical noise in the FIB tool must be reduced below that in order to make the measurements. This is both technically difficult and quite costly.

4.4.2.2 Gallium implantation

The residual gallium that has been implanted or embedded during imaging and micromachining is detrimental to repair of phase shift masks. Ga-implanted quartz has greatly reduced optical transmissivity. For repair of masks exposed to deep ultraviolet (DUV) sources, gallium probably cannot be used due to reduction in transmission through gallium-implanted quartz.

There is serious concern regarding gallium implanted in circuits during device modification or in-line process monitoring. The effects of gallium ions on IC performance have not been quantified. Therefore there is as much incentive to identify alternative ion-source technology as there is to pursue a reduction in beam size.

4.4.3 Beyond Gallium (Liquid Metal Ion Source) Gas Field Source

The chief drawback to an LMIS is the M, i.e., the metal ion. Different source materials have been tried, such as Be, Si, and Au with limited success due to (1) the basic instability of ion emission, and (2) beam current density being too low, which governs the overall writing speed.

An alternative approach is to focus a beam of nonmetal, inert gaseous ions, such as hydrogen, helium, or neon through charged particle optics. This gas field ion source (GFIS) approach would avoid the disadvantages of metal ions, such as staining of optically transmissive materials, and the introduction of mobile metal ions into electrostatic circuits during micromachining. An even more attractive advantage is the very high beam current densities that could be theoretically achieved. Researchers at Cornell University[32] and the Max Planck Institute[33] have obtained ion beams with a brightness of almost 10 A/cm^2/sr, which is 10 to 100 times greater than can be obtained with an LMIS.

At present, GFIS technology has not been commercially merged with FIB products. A successful GFIS system requires stringent integration of an ultrahigh vacuum environment, high voltage, and ultrapure gas handling systems with standard FIB technology before commercialization is possible[34].

REFERENCES

1. T. D. Cambria and N. P. Economou, "Mask and Circuit Repair with Focused-Ion Beams," Solid State Technol. **30**, 133 (1987).

2. J. Melngailis, R. T. Post, M. W. Geis and R. W. Mountain, "The Focused Ion Beam As Integrated Circuit Restructuring Tool," J. Vac. Sci. Tech. **B4**, 176 (1986).

3. R. G. Lee and J. C. Morgan, "Integration of a Focused Ion Beam System in a Failure Analysis Environment," Proc. ISTFA '91 (International Symposium for Testing and Failure Analysis) **85** (1992).

4. R. J. Young, E. C. G. Kirk, D. A. Williams, and H. Ahmed, "Fabrication of Planar and Cross-Sectional TEM Specimens using a Focused Ion Beam," Mat. Res. Soc. Symp. Proc. **199**, 205 (1990).

5. M. J. Vasile, C. Biddick, and S. Schwalm, "Microfabrication by Ion Milling: The Lathe Technique," J. Vac. Sci. Tech. **B 12** (4), 2388 (1994).

6. W. B. Thompson and R. G. Lee, "Focused Ion Beam Process Monitoring," Proc. SPIE **1673**, 414 (1992).

7. P. Blauner and J. Mauer, "X-ray Mask Repair," IBM J. Res. Development **37** (3), 421 (1993).

8. J. D. Casey, Jr., A. F. Doyle, R. G. Lee, and D. K. Stewart, "Gas-Assisted Etching with Focused Ion Beam Technology," Microelectron. Eng. **24**, 43 (1994).

9. M. Abramo, L. Hahn, and L. Moskowicz, "Gas Assisted Etching: An Advanced Technique for Focused Ion Beam Device Modification," Proc. ISTFA **439** (1994).

10. D. M. Thaus, T. J. Stark, D.P. Griffis, and P. E. Rusell, "Development of Foucsed Ion-Beam Machining Techniques for Permalloy Structures," J. Vac. Sci. Tech. **B6**, 3928 (1996).

11. A. Dubner and A. Wagner "The Role Of Gas Absorption In Ion-Beam Induced Deposition Of Gold," J. Appl. Phys. **66**, 870 (1989).

12. J. R. Cleaver, H. Ahmed, P. J. Heard, P. D. Prewitt, G. J. Dunn and H. Kaufamn, "FIB Repair Techniques for Clear and Opaque Defects in Masks," Microelectron. Eng **3**, 253 (1985); **6**, 617 (1987).

13. J. P. Levin, P. G. Blauner, and A. Wagner, "Model for Focused Ion Beam Deposition", Proc. SPIE **1263**, 2 (1990).

14. L. R. Harriott and M. J. Vasile, "Focused Ion Beam Induced Deposition of Opaque Carbon Films," J. Vac. Sci. Tech., **B6**, 1035 (1988).

15. J. S. Ro, A. D. Dubner, C. V. Thompson, and J. Melngailis, Mat. Res. Soc. Symp. Proc. **101**, 255 (1988).

16. T. Tao, J. S. Ro, and J. Melngailis, "Focused Ion Beam Induced Deposition of Platinum," J. Vac. Sci. Tech. **B8**. 1826 (1989).

17. K. Gamo, D. Takehara, Y. Hamamura, M. Tomita, and S. Hamba, "Maskless Ion Beam Assisted Deposition of W and Ta," Microelectron. Eng. **5**, 163 (1986).

18. W. P. Robinson, "The Application of Ion Beam Assisted Deposition of Chrome to Photomask Repair," Proc. SPIE **1089**, 228 (1989).

19. M. E. Gross, L. R. Harriott, and R. L. Opila, Jr., "Focused Ion Beam Stimulated Deposition of Aluminum from Trialkylamine Alanes," J Appl. Phys. **68**, 4820 (1990).

20. H. Nakamura, H. Komano, K. Norimatu, and Y. Gomei, "Silicon Oxide Deposition into a Hole Using a Focused Ion Beam," Jpn. J. Appl. Phys. **30** (11B) 3238 (1991).

21. M. Abramo and L. Hahn, "Ion Beam Induced Insulator Deposition for Chip Circuit Modification," Proc. International Conference on Metallurgical Coatings and Thin Films (ICMCTF) 1995.

22. D. K. Stewart, L. A. Stern, and J. C. Morgan, "Focused Ion Beam Induced Deposition of Metal for Microcircuit Modification," Proc. SPIE **1089**, 18 (1989).

23. A. Wagner, J. P. Levin, J. L. Mauer, P. G. Blauner, S. J. Kirch, and P. Longo, "X-ray Mask Repair with Focused Ion Beams," J. Vac. Sci. Technol. **B8** (6), 1557 (1990).

24. Semiconductor Industry Association "The National Technology Roadmap for Semiconductors," 1994 **SIA,** 4300 Stevens Creek Blvd., Suite 271, San Jose, CA 95129.

25. J. A. Lange and S. Czapski, "Focused Ion Beam Applications for Design and Product Analysis," Proc. ISTFA, **397** (1991).

26. T. A. Core, W. K. Tsang, and S. J. Sherman, "Fabrication Technology for an Integrated Surface-Micromachined Sensor," Solid State Tech. **36** (10) 39 (1993).

27. R. Hill, Micrion, private communication.

28. G. L. R. Mair and T. Mulvey, "Ion Beam Lithography," Microelectron. Eng. **3**, 133 (1985).

29. Y. Chen, R. J. Nicholas, G. M. Sundaram, P. J. Heard, P. D. Prewitt, J. E. Frost, G. A. C. Jones, D. C. Peacock, and D. A. Ritchie, "Low-field Magneto Transport Study of Localization in a Mesoscopic Antidot Array," Amr. Phys. Soc.**47**, 7354 (1993).

30. A. Wagner, P. Longo, S. Cohen, and P. Blauner, "Focused Ion Beam Metrology," page presented at Symposium on Electron, Ion and Photon (EIPB) Beams, (1995).

31. T. J. Stark, G. M. Shedd, J. Vitarelli, D. P. Griffis, and P. E. Russell, "H_2O Enhanced Focused Ion Beam Micromachining," Proc. EIPB (1995).

32. P. R. Schwoebel and G. R. Hanson, "Beam Current Stability from Localized Emission Sites in a Field Ion Source," J. Vac. Sci. Tech. **B3**, 214 (1985).

33. K. Jouesten, K. Bohringer, and S. Kalbitzer, "Current-Voltage Characteristics of a Gas Field Ion Source," Appl. Phys. **B46**, 313 (1988).

34. W. Thompson, A. Armstrong, S. Etchine, R. Hill, S. Kalbitzer, R. Percival, A. Saxonis, and C. Wilbertz, "Towards a Commercial Gas Field Ion Source," Proc. SPIE **2437**, 284 (1995).

CHAPTER 5

Plating Techniques

Lubomyr T. Romankiw
Eugene J. M. O'Sullivan
IBM Corporation
Thomas J. Watson Research Center

CONTENTS

5.1 Introduction / 199
  - 5.1.1 Electroplating / 199
  - 5.1.2 Electrode potential / 201
  - 5.1.3 Single element plating—not all metals can be electrodeposited from an aqueous electrolyte /204
  - 5.1.4 Alloy plating from aqueous solutions / 207

5.2 Electroless Deposition / 208
  - 5.2.1 Some basic aspects of electroless deposition / 208
  - 5.2.2 Electroless nickel / 210
  - 5.2.3 Electroless copper / 211
  - 5.2.4 Electroless gold / 212

5.3 Aspects of the Mechanism of Electrodeposit Formation / 212
  - 5.3.1 Nucleation and growth of deposits / 212
  - 5.3.2 The effect of current density / 214
  - 5.3.3 Aspects of the structure of Electrodeposits / 216
    - 5.3.3.1 Epitaxial growth and preferred orientation / 217
    - 5.3.3.2 Preferred orientation of grains / 217
  - 5.3.4 Properties of Electrodeposits / 219
  - 5.3.5 Stress in plated films / 220
  - 5.3.6 Other properties / 222
  - 5.3.7 Additives and plating formulations / 223
  - 5.3.8 Brighteners / 224
  - 5.3.9 Levelers / 225
  - 5.3.10 Stress reducing agents / 225

- 5.4 Current Density Uniformity / 227
  - 5.4.1 The various types of current density distribution / 227
    - 5.4.1.1 Primary current distribution / 229
    - 5.4.1.2 Secondary current distribution / 230
    - 5.4.1.3 Tertiary current distribution / 230
  - 5.4.2 Diffusion layer / 231
  - 5.4.3 Uniformity of current density across a substrate / 233
  - 5.4.4 Local current density distribution / 235
    - 5.4.4.1 Resist pattern-scale uniformity / 235
    - 5.4.4.2 Feature-scale uniformity / 236
- 5.5 Electroplating in the Electronics Industry / 237
  - 5.5.1 Significance of plating-through-mask technology / 237
  - 5.5.2 Impact of plating-through-mask technology / 239
  - 5.5.3 Plating through lithographic masks—process details / 239
  - 5.5.4 Electroforming in electronics is known by many names / 242
  - 5.5.5 Patterning and pattern transfer / 243
  - 5.5.6 Beginnings of plating-through-mask technology in electronics / 247
  - 5.5.7 Bubble memory, e-beam, and x-ray lithography / 252
  - 5.5.8 Computer on a chip / 254
  - 5.5.9 Multi-chip thin film package / 255
  - 5.5.10 High precision and tight tolerances—new equipment needed / 256
- 5.6 LIGA and HI-MEMS / 258
  - 5.6.1 Single layer structure / 258
  - 5.6.2 Sacrificial layers and release of electroplated components / 259
- 5.7 Examples of Metal Electrodeposits / 264
- 5.8 Fabrication of Magnetic Minimotors / 270
  - 5.8.1 Fabrication overview—optical / 271
  - 5.8.2 Lithography—optical / 273
  - 5.8.3 Permalloy electrodeposition / 274
  - 5.8.4 Copper electrodeposition / 274
  - 5.8.5 Dielectric materials and planarization—optical / 275
  - 5.8.6 Completion of the minimotor structure / 275
  - 5.8.7 X-ray minimotor fabrication / 276
    - 5.8.7.1 Process compatibility with x-ray lithography / 279
- 5.9 Discussion / 280
  - 5.9.1 Minimum requirements to be able to plate through the mask or practice LIGA / 280
  - 5.9.2 Resist residue / 283
  - 5.9.3 Plating into deep recesses—current and mass distribution / 283
- 5.10 Summary / 287

Suggested Further Reading / 288
Acknowledgments / 289
References / 290

5.1 INTRODUCTION

5.1.1 Electroplating

Electroplating is a process for metal and/or metal alloy deposition by electrolysis from a salt solution. A metal salt dissolved in a polar solvent is present in an ionic form; the most commonly used solvent is water. There are a great number of metals and metal alloys that can be electrodeposited from an aqueous solution. Plating from aqueous solutions is commercially used in decorative metal applications; in the automotive, aerospace, and electronic industries; in metal recovery from ores, and many other applications. Most of the electroplating from aqueous solutions is carried out in a temperature range of 10 to 100°C.

In electronic device and microelectromechanical systems (MEMS) fabrication, it is preferable to electroplate at or near room temperature. This is primarily because of the need for compatibility with organic resists and other materials that may be part of the structure.

Metals and metal alloys can also be plated from polar organic solvent-based solutions and from molten salts. It is primarily the refractory metals that are plated from molten salts. On a large-scale commercial basis, such molten salt-based electrolysis is used to recover aluminum from its ores. Because of the high temperatures and frequently very aggressive nature of the molten salts, they cannot be used for electronic device or MEMS device fabrication. The polar organic solvent electrolysis of salt has not been sufficiently extensively investigated to be commercially useful in either electronics or MEMS.

The most frequently electroplated metals in electronics include Cu, NiFe, NiFeCo, SnPb, NiP, AuNi, AuCo, Pd, and PdNi. Copper is used primarily as a conductor; NiFe, NiFeCo, and other combinations with transition metal alloys serve as magnetic materials in write heads and CoPt, CoNiP, and NiP in media; Sn and SnPb are used as solders and solder interconnects between the chips, chip carriers and PC boards; and Ni, NiP, Au, Pd, AuNi, AuCo, and PdNi find application in connectors.

Electrodeposition through a resist mask is a basic fabrication procedure for electronic devices and for 3D structures in the special area of MEMS known as HI-MEMS, which deals with thick structures that are often several hundred microns thick. The technique uses selective electroforming of structures on conductive seed layers. Physical vapor deposition techniques, such as evaporation or sputtering, are less suitable for building thick, high aspect ratio structures because of the lower deposition rates and the inability to fill features selectively. If the metal is deposited as a blanket layer by evaporation, sputtering, plating, or other means and is then patterned by chemical etching, it is difficult to control the pattern within tight tolerances because of undercutting and roughening of the features. It is also impossible to obtain high aspect ratio structures by chemical etching.[1–3] Patterning by sputter etching, ion milling, and lift-off also have disadvantages compared with deposition through a photoresist mask.[1–3]

The electroplating of a variety of metals and metal alloys from aqueous solutions has been extensively investigated for many decades. A number of textbooks and review articles may be consulted for an introduction to electrodeposition fundamentals and practice.[4,5] The important issues in electrodeposition include electrode kinetics, mass transport, cell design, and the structure of electrodeposits. The formulation of the solution is of immense importance and affects these aspects of electrodeposition to varying degrees. In general, these subareas mutually influence one another and cannot usually be studied independently. The initial stages of electrodeposition have been discussed in depth in a number of books.[4–7] This chapter gives a general overview of electrodeposition, discusses electroplating through-mask technology, and gives selected examples of the fabrication of some commercially marketed electronic devices and experimental MEMS devices.

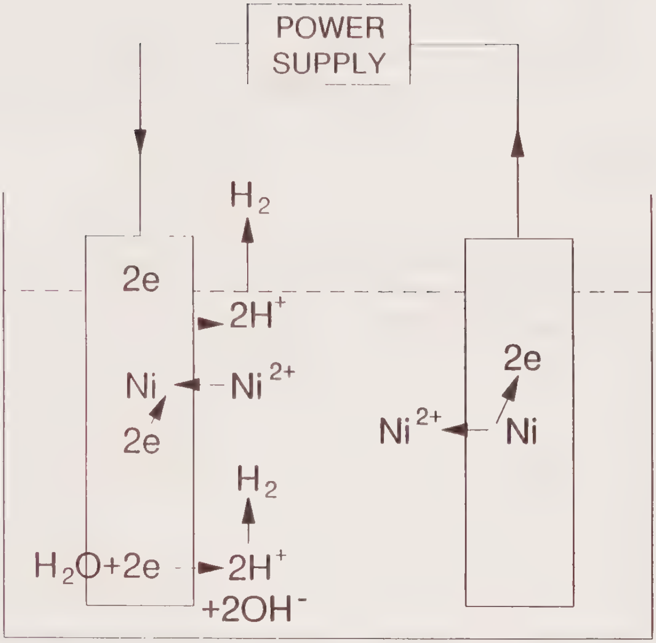


Fig. 5.1. Schematic representation of an electrochemical cell showing Ni electrodeposition and H_2 evolution at the cathode and Ni dissolution at the anode.

5.1.2 Electrode Potential

In a typical electroplating setup, shown in Fig. 5.1, two electrodes are immersed in a plating solution. One of the electrodes serves as an anode and the second as a cathode. As the current is supplied by an external power supply, the negatively charged ions (anions) flow to the anode, where they pair up with the cations formed by dissolution of the soluble electrode; the positively charged ions flow to the cathode where they acquire electrons and deposit in the form of a metal film.

When a metallic electrode is inserted into an aqueous solution containing cations of that metal, an equilibrium is established between the metal and the solution according to the equation:

$$M^{o} \rightleftarrows M^{n+} + ne^{-} \qquad (5.1)$$

where M^o represents the metal, M^{n+} represents the metal ion, n is the valence of the ion, and ne represents the number of electrons released. The equation represents the equilibrium between the tendency of the latter to enter the solution as cations, and the tendency of the metal to deposit out of the solution onto the electrode as a zerovalent metal. When no external current is applied, the metal substrate will become negatively or positively charged, depending on whether the forward or the reverse reaction predominates.

To define the potential between the solution and the electrode it is essential to have a reference point; electrochemists have chosen the hydrogen equilibrium reaction as such a reference point. This reference point is the potential established when hydrogen is bubbled past a platinized platinum (inert) electrode immersed in an acid electrolyte. The equilibrium reaction shown in equation (5.2) is defined as the reference potential, or $E^o_{H_2}$, which is assigned a value of 0.0 V:

$$H_2 \overset{\leftarrow}{\rightarrow} 2H^+ + 2e^- \tag{5.2}$$

A similar potential exists between oxygen gas and water, which is 1.23 V positive of the hydrogen electrode:

$$O_2 + 4H^+ + 4e^- \overset{\leftarrow}{\rightarrow} 2H_2O \tag{5.3}$$

The forward and backward rates of the oxygen electrode are extremely slow relative to the hydrogen electrode, so that the potential of 1.23 V is virtually unattainable in practice.

The potential between a metal and a solution of its ions is expressed by the Nernst equation:

$$E = E^o_M + \frac{RT}{nF} \ln \frac{a}{a_M} \tag{5.4}$$

where E = the potential in volts, E^o_M = a standard electrode potential for the given metal with reference to the hydrogen equilibrium potential $E^o_{H_2}$; R = the gas constant, 8.314 $J \cdot K^{-1} mole^{-1}$; T = absolute temperature in kelvins; F = the Faraday constant, 96,500 C/g-equiv. (the gram equivalent is obtained by dividing the gram atomic weight of the metal by the number of electrons received per atom during deposition); n = valence; a = the activity of metal ion in solution; and a_M = the activity of the solid metal (which in the case of pure a solid metal is equal to 1.0). In a highly ionized, dilute plating solution, activity and the concentration C_b of the metal ion in mol/liter are nearly equal to each other. Therefore a is usually replaced by C_b, so that equation (5.4) simplifies to:

$$E = E_M^o + \frac{RT}{nF} \ln C_b \tag{5.5}$$

where C_b is the bulk concentration of the metal ion in the electrolyte.

Table 5.1: Selected standard electrode potentials in aqueous solutions at 25°C.

Reaction	$E_{25^{\circ}C}$ (V vs. NHE[a])
$Li^+ + e^- \rightarrow Li$	-3.010
$Ca^{2+} + 2e^- \rightarrow Ca$	-2.860
$Mg^{2+} + 2e^- \rightarrow Mg$	-2.360
$Ti^{2+} + 2e^- \rightarrow Ti$	-1.750
$Al^{3+} + 3e^- \rightarrow Al$	-1.662
$Zn^{2+} + 2e^- \rightarrow Zn$	-0.762
$Cr^{3+} + 3e^- \rightarrow Cr$	-0.744
$Fe^{2+} + 2e^- \rightarrow Fe$	-0.447
$Cd^{2+} + 2e^- \rightarrow Cd$	-0.403
$Co^{2+} + 2e^- \rightarrow Co$	-0.280
$Ni^{2+} + 2e^- \rightarrow Ni$	-0.257
$Mo^{3+} + 3e^- \rightarrow Mo$	-0.200
$Sn^{2+} + 2e^- \rightarrow Sn$	-0.137
$Pb_{2+} + 2e^- \rightarrow Pd$	-0.126
$H^+ + e^- \rightarrow \frac{1}{2}H_2$	0.000
$Cu^{2+} + 2e^- \rightarrow Cu$	+0.342
$PdCl_4^{2-} + 2e^- \rightarrow Pd + 4Cl^-$	+0.591
$Ag^+ + e^- \rightarrow Ag$	+0.799
$Au^+ + e^- \rightarrow Au$	+1.692

The standard electrode potentials of the elements are arranged in a series known as the electromotive series (in aqueous solution). Similar electromotive series can be constructed for each polar organic electrolyte, and for each molten salt electrolyte. These latter series, however, are neither as well developed nor as extensively studied as the electromotive series for aqueous solutions. Table 5.1 lists selected standard electrode potentials in aqueous solutions at 25°C.

---

[a] Normal hydrogen electrode

5.1.3 Single-Element Plating — Not All Metals Can Be Electrodeposited from an Aqueous Electrolyte

Typical reactions that take place on the cathode during the deposition of metals having a standard electrode potential negative of the hydrogen electrode are as follows:

$$M^{+n} + ne^- \rightleftarrows M^o \quad \text{Metal deposition} \tag{5.6}$$

$$2H^+ + 2e^- \rightleftarrows H_2\uparrow \quad \text{Hydrogenevolution(H.E.)(inlowpHsoln.)} \tag{5.7}$$

$$2H_2O + 2e^- \rightleftarrows H_2\uparrow + 2OH^- \quad \text{H.E. (in neutral/basic soln.)} \tag{5.8}$$

The more negative the standard electrode potential, the more difficult it is to deposit the indicated metal because the competing reactions are increasingly favored. Reactions that have a more negative standard electrode potential than hydrogen deposit with the evolution of hydrogen as a side reaction. Electrodeposition of most of the non-noble metals such as Ni, Fe, Co, and Zn is accompanied by H_2 evolution.

The highly reactive and/or refractory metals Li, Ca, Mg, Ti, and Al shown above the upper line in Table 5.1 cannot be electrodeposited from aqueous solutions. Their standard potentials are so negative that aqueous solutions containing these ions yield H_2 gas at the cathode at low current densities, and yield H_2 plus metal hydroxides at higher current densities. This happens because water tends to decompose according to equation (5.8) before metal ion reduction can take place according to equation (5.6).

The electrodeposition of metals from aqueous solutions of simple salts in the acidic range of pH values becomes possible starting with Zn and metals below it in Table 5.1. In principle it should not be possible to electroplate any metal above the hydrogen gas formation reactions at $E_{25°C} = 0.00$ V, (equations (5.7) or (5.8)).

It is rather fortuitous that Nature provides us with a significant hydrogen overpotential for many metals and alloys. As the potential becomes more negative, water may begin to decompose on the cathode according to equation (5.8), resulting in not only evolution of H_2 but also formation of OH^- in solution in the vicinity of the cathode. It takes a certain amount of energy in the form of a potential that has to be applied to the cathode to overcome the energy of activation for hydrogen evolution and hydrogen atom recombination into H_2 molecules, which then nucleate into gas bubbles. The difference between this potential and the standard potential for the hydrogen electrode reaction is known as the *hydrogen overvoltage,* or *overpotential.* Figure 5.2 is a schematic representation of this effect.

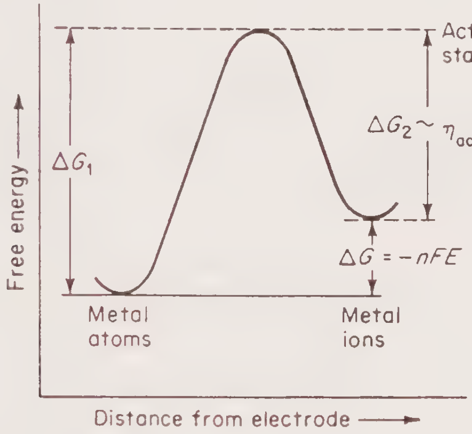


Fig. 5.2. The energy barrier in metal deposition.

The hydrogen overpotential is different on each metal. It is the hydrogen overvoltage which makes possible electrodeposition of Pb, Sn, Mo, Ni, and other metals all the way to Zn. Since deposition of these metals according to equation (5.6) is always accompanied by hydrogen evolution according to equations (5.7) and (5.8), the current efficiency during electrodeposition is less than 100%. Typically it may range from 60 to 95%, the remaining current going to hydrogen evolution from protons or water molecules.

The metals with more positive standard potential than copper — such as Pd, Ag, Au, and Pt, are considered to be "noble." The noble metals are not usually electroplated from their simple salt solutions. These solutions are not very stable; metal may precipitate out of solution. It is usual to formulate electroplating solutions of the noble elements using some form of complexation to maintain the metal in solution and to increase the overpotential for electrodeposition. However, it takes a certain amount of activation energy to decompose the complex; in general, the metal ion must not be so strongly stabilized by the complexant that it will not undergo electroreduction. Therefore, electroplating of the noble metals from a complex in solution is also accompanied by hydrogen evolution. This means that the electrodeposition process for these metals is also less than 100% efficient. Only copper, which has a positive standard potential and can be used as a stable, simple salt solution such as $CuSO_4$ in H_2SO_4, can be deposited with a current efficiency of 100%.

One of the most important complexants is cyanide, which is widely used in gold plating. Most cyanide containing solutions need to be operated at a relatively alkaline pH, although the cyano-gold complex is stable at pH values as low as 3. Because of very high toxicity of cyanide, when using the cyanide complex-based plating solutions, precaution should be taken not to contact the solution, not get the solution in contact with acids, not to rinse the parts with acid, and to dispose of both the bath and the rinsing solution according to U.S. Occupational Safety and Health Administration rules.

Often complexing of the non-noble metals is also carried out to permit electroplating over a wide range of pH values. The non-noble metals in near neutral and alkaline pH values tend to hydrolyze and precipitate out of the aqueous solution as hydroxides. The common complexant used for transition metal elements such as nickel, iron, and cobalt plating baths is sulfamate. It is frequently used in the case of Ni baths.[5] Electroplating solutions are formulated with various other complexants, such as carboxylic acids, i.e., tartate, citrate, maleate, and others. These acids are usually used in salt form for the convenience of pH adjustment. These complexing acids ionize at pH values greater than ~4 and form mono, di, or even trivalent anions as the pH is increased. Since it is the anion that ties up the metal ion, the strength of the complex for each anion increases as the pH increases. Each anion has a different pH range in which it is effective.

Electrodeposition is governed by the laws of electrolysis, where the current-time relation may be used to produce a defined mass or thickness of deposit, assuming that side reactions such as H_2 evolution are taken into account. The efficiency of 100% observed for acid Cu baths is generally the exception rather than the rule for other metals. The Faraday law states that the weight of the metal deposited is given by the equation:

$$g = \frac{I\,e\,t}{F} = \frac{I\,e\,t}{96{,}500} \tag{5.9}$$

where g = metal deposited in grams, I = current in amperes, e = chemical equivalent weight, t = time in seconds the current is passed, and F = Faraday constant. The *equivalent weight* is calculated by dividing the molecular weight of the metal by the valence (the number of electrons it must acquire at the cathode to deposit as a metal).

In most of the practical cases, because H_2 is evolved according to equations (5.7) and (5.8), the actual amount of metal deposited is less than is calculated from equation (5.9). If the actual weight is measured and compared with the value calculated from equation (5.9), one arrives at the efficiency of the system.

The current efficiency depends on the pH of the solution; the current density; the temperature, amount, and type of organic additive; and in case of complexed ions, also on the strength of the ion complex. An additional problem that affects the accurate prediction of deposit thickness is the estimation of the actual, local, current density. Since current takes the path of least resistance, it may not be uniformly distributed over the surface of the part being plated, as discussed later. This is especially the case if electroplating is done through a polymeric resist mask that has patterns greatly differing in sizes, shapes, and densities.

5.1.4 Alloy Plating from Aqueous Solutions

It is possible to electroplate many binary, tertiary, and even quaternary metal alloys. While one would expect the ratio of the metals in the deposit to be the same as the ratio of metal ions in the simple salt solution, very few binary alloy plating systems even approximate this condition. Most alloys, when plated from simple salt solutions, deviate considerably from the composition of the electrolyte.

One of the most striking examples of this is the Ni and Fe alloy (Permalloy). It is necessary to have a ratio of Ni to Fe in a simple salt solution as high as 80:1 to achieve a Permalloy electrodeposit with a Ni:Fe ratio of 4:1, or 20% Fe, 80% Ni. This phenomenon is often referred to as *anomalous codeposition.* Similar anomalies exist between Co and Ni, Zn and Ni, and to a lesser degree between Sn and Pb and other metal pairs.

The NiFe system has been extensively studied in connection with the electrodeposition of thin film recording heads. The most frequently used reference is that of Dahms and Croll.[8] These authors provided the first in-depth explanation of the mechanism of anomalous codeposition. The reader is also referred to a chapter by Andricacos and Romankiw[9] for further discussion of anomalous codeposition. These authors also treat the micromagnetic requirements for thin film head applications and the through-mask electroplating process for copper and Permalloy, describe the fabrication of thin film heads, and review a large number of Permalloy electroplating baths which appear in the published literature. They also discuss the effect of individual plating parameters on magnetic properties.

Magnetic properties are strongly affected by the structure of the deposit. One of the first review papers that demonstrated a close correlation of the magnetic properties and structure of electrodeposits was that of Romankiw and Thompson.[10] For more recent work on the mechanism of Permalloy deposition, the reader is referred to papers by Deligianni and Romankiw[11]; Hessami and Tobias[12]; Sasaki and Talbot[13]; and Yin, Papov and White.[14]

5.2 ELECTROLESS DEPOSITION

5.2.1 Some Basic Aspects of Electroless Deposition

Electroless deposition may be defined as the deposition of a metallic coating through controlled chemical reduction by a reducing agent that is catalyzed by the metal or alloy being deposited. No external power supply is required. Electroless plating plays a major role in metal coating of parts for automotive, aerospace, electronic, and household appliances. The metal salts and reducing agent react in the presence of a suitable catalyst, which is either the surface being plated, or a catalyst deposited on it. Chemical reducing agents usually employed are sodium hypophosphite and dimethylamine borane in the case of Ni and Ni alloys, and formaldehyde in the case of Cu. Other less widely used reducing agents include hydrazine and sodium borohydride. The metals that can be deposited by electroless means are Cu, Ag, Ni as Ni(P) or Ni(B); Co as Co(P) or Co(B), CoNi(P) or CoNi(B); NiFe(P); Au, Pd and several alloys of Ni with different refractory metals.

The substrate metal may be catalytically active for electroless deposition if properly cleaned, for example, Cu in the case of baths containing formaldehyde and dimethylamine borane reducing agents. Insulators and metals that are not catalytically active for initiation of electroless deposition are usually catalyzed, or "activated" by deposition of a Pd catalyst. In the case of insulators, a colloidal-type Pd catalyst is adsorbed on the surface to initiate the plating process, which is then carried on by the newly deposited metal. Often in the case of catalytically inactive metals it is possible to deposit a thin layer of dispersed Pd nuclei on the less noble metal surface by a chemical displacement process in an acidic solution. In this process, a small amount of the substrate metal dissolves and is stoichiometrically replaced by the Pd catalyst. No reducing agent is involved. More complicated forms of catalyzation are required for a metal such as Al and its alloys due to the ever-present surface oxide.

A notable feature of electroless deposition is that it is in principle easier to obtain conformal coatings using this process than using the electrodeposition process discussed below. This behavior is possible with electroless solutions since they often operate under conditions of significant kinetic control of deposition. However, the presence of small amounts of additives such as Pb (typically 1-2 ppm) tends to cause non-planar diffusion at small isolated features and edges of lines so that these tend to plate at a slower rate than average. Such strongly adsorbing "poisons" are added to prevent spontaneous formation of metal nuclei in solution, and thus avoid catastrophic bath plate-out.

It is the nature of electroless plating baths that they are operated at the verge of instability. They must be active enough to deposit metal on a catalytic surface, but not so active that they would start homogeneous nucleation in solution and a catastrophic bath decomposition. The proper formulation of the electroless solution is extremely critical. Because electroless deposition is surface catalyzed, it is extremely strongly affected by even the smallest traces of impurities, especially if they have a surface-active nature.

Because of its selectivity and the fact that electrically isolated features can be easily plated, electroless deposition is a powerful technique. It does have some disadvantages, however: (a) only a limited number of metals and alloys can be deposited by electroless methods compared with electroplating; (b) a pure metal is obtained in only a few cases: e.g., pure Cu is obtained in the presence of a formaldehyde reducing agent, while an Ni(P) alloy is obtained in the case of electroless Ni solutions containing hypophosphite as a reducing agent; (c) deposition rates are generally much lower for electroless solutions than for electrodeposition, e.g., 3-5 μm/hr for Cu/formaldehyde baths; and (d) a well-behaved activation process and careful control of the additive is necessary in order to properly initiate plating, and to achieve good uniformity in the case of plating features whose size approaches the diffusion layer thickness for the additive.

5.2.2 Electroless Nickel

Electroless nickel plating is extensively used in the household appliance, automotive and electronic industries. Electroless nickel is plated using sodium hypophosphite or dimethylamine borane as a reducing agent. During the reduction process, phosphorus and boron are codeposited with Ni. These two elements impart some unusual and very desirable properties to electroless nickel; e.g., good corrosion resistance, high resistance to wear, hardness and some others. In the household appliance and automotive industries, electroless NiP and electroless NiB are used to produce corrosion-protective layers in place of chrome. A very extensive use of electroless NiP and NiB is made in the plating of plastic parts that replace solid metal components, cutting cost while still achieving durable, mechanically hard, and wear-resistant surfaces. Electroless NiP is used extensively to overcoat aluminum disks to produce a light but very hard and durable nonmagnetic substrate for magnetic disk fabrication. Today almost all magnetic disks around the world are prepared by using AlMg disks coated with electroless NiP. The AlMg surface requires special preparation and activation which involves thorough cleaning of Al oxide, and zincating (immersion, or chemical exchange plating of AlMg with Zn) in order to initiate NiP electroless plating.

Nickel phosphorus and NiB are also extensively used in plating ceramic chip carriers to provide a metal transition bridge from Mo- or W-fired (sintered) paste, as well as to act as a diffusion barrier between copper and SnPb. Nickel phosphorus is receiving an exceptional amount of attention because of its unusual properties which cannot be achieved in any other way. These are (a), the ability to plate selectively not only on metals but also on plastics, ceramics, glass, and other nonconducting materials; (b) deposition in an amorphous form that subsequently can be dispersion hardened by thermal treatment to produce Ni_3P dispersed in an Ni matrix, with unusually high resistance to wear and high hardness. There is a considerable literature available covering various aspects of this material. The reader is referred to good book references by Reidel[15] and by Hajdu.[16]

5.2.3 Electroless Copper

Electroless copper is usually plated from an ethylenediamine tetraacetate copper complex in an alkali bath using formaldehyde as a reducing agent. It plays an extremely significant role in electronics in two-sided, multilevel printed board fabrication.

Because of its ability to uniformly coat the inside of very small holes, it is used to connect the two sides of double-sided PC boards. It is also used extensively to plate very narrow and long vias in multilevel boards in which the copper deposited in the holes connects not only the two sides of the board, but as many as 40 to 60 levels of inner planes. The understanding and control of both the plating bath and the metal structure of the electroless copper have been perfected to the level that this process is now used in production of complete multilevel boards and flex circuits by selective electroless plating. It has become an enabling technology in second-level PC board packaging. There are many excellent review papers on electroless copper[17-19] and books discussing the uses of copper in PC boards and other electronic applications.[20-24]

Unlike electroless Ni, copper is plated without incorporating the fragments of the reducing agent or the complexants. Thus pure copper is obtained whose grain size and orientation can be controlled to be equiaxial. This gives electroless copper good ductility, elongation and electrical conductivity. The grain size and orientation are controlled by adding trace amounts of cyanide, which reduces the tendency for the deposit to grow in a columnar fashion, and thus be weak in tension and crack during soldering.

The various advantages enumerated above, including the ability to plate into deep recesses and blind vias, far outweigh the disadvantage that electroless baths plate at rather low rates.

5.2.4 Electroless Gold

Electroless gold is receiving attention and growing in importance in the area of contacts and connectors for reasons of low contact resistance, excellent corrosion resistance, and the selectivity of plating of precious metal which translates into cost savings. Electroless gold, which once was plated exclusively from cyanide plating solutions using borohydride as a reducing agent, suffered from the disadvantages of the toxicity of cyanide and the potential explosiveness of borohydride. Present-day solutions are much less toxic and contain safer reducing agents, such as dimethylamine borane. This improvement, in addition to the advantage of selectivity, is making electroless gold an important electroless plating process. Ganu and Mahapatra discuss the uses of electroless gold in the electronic industry[25] and Ali and Christie review the electroless gold deposition process,[26] while Okinaka and Wolowodiuk,[27] and Canaperi, Jagannathan and Krishnan[28] discuss the replenishment schemes for electroless gold solutions. The number of turnovers of the plating solution and replenishment schemes are important issues in the use and maintenance of electroless gold baths. A practical solution to this problem was proposed by Canaperi, Jagannathan, and Krishnan.[28]

5.3 ASPECTS OF THE MECHANISM OF ELECTRODEPOSIT FORMATION

5.3.1 Nucleation and Growth of Deposits

The current density is a key parameter in electroplating. We will discuss later how the current is distributed across the cathode. Local current density determines the film thickness and many intrinsic film properties. To understand how current density and the plating solution influence the deposit it is best to have an understanding about the mechanism of deposit formation.

In general, deposits from practical plating solutions are crystalline. At the beginning of the plating process, deposition may occur on an electrode surface of a different material, e.g., Permalloy on a Cu surface. Hence the first step will be the formation of nuclei of the new phase and their growth into crystals with a lattice structure characteristic of the material being electroplated. Deposit nucleation is always an improbable event, and for most metals it is achieved at an electrode surface under conditions of large overpotential. This is because the formation of the new phase is always difficult, in large part due to the highly energetic nature of small particles resulting from their high surface-to-volume ratio. Even though the bulk free energy change for the electrodeposit phase may be negative and the deposit thermodynamically stable, the surface energy contribution is normally positive and dominant for deposit nuclei. The various surface energy terms involving the electrodeposit, electrode, and solution are only overcome for nuclei with a small radius by applying a high overpotential.[29]

The density of nucleation is dependent on the composition of the plating solution and the current density. Once formed at high overpotentials, nuclei continue to grow quite rapidly at low overpotentials, and in a constant-current situation — usually the case in practical electroplating — the potential will decrease substantially once nucleation has been achieved.

Crystal growth occurs by incorporation of the individual atoms into the lattice. Newly deposited atoms may be expected to be stable only when they enter the lattice structure at a site where they strongly interact with several other atoms already in the lattice. Generally, incorporation will be most favorable at a kink site (where an atom can interact with three neighbors), while some incorporation may also occur at edge sites (two atoms available for interaction) as can be seen in Fig. 5.3.

The structure of the growing layer will be substantially determined by the rates of electron transfer to form the adatom, diffusion of the adatom across the surface into a lattice position, and the current density. Surface diffusion will be fast compared with electron transfer at low current density, and the newly deposited atom is likely to end up in a favored

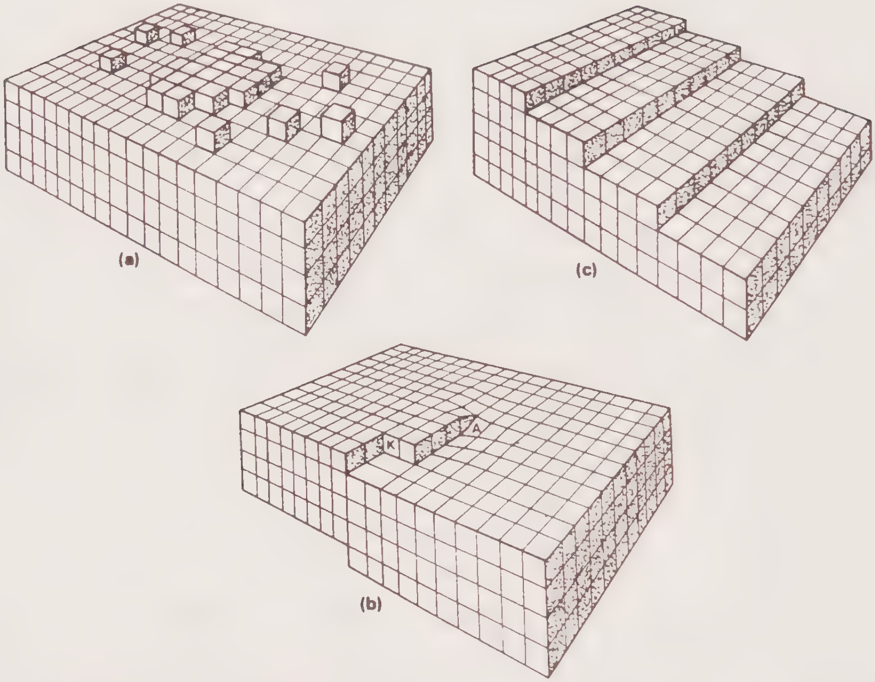


Fig. 5.3. Models of the different kinds of microsteps on a surface: (a) two-dimensional nucleus, (b) emergent screw dislocation (A) with kink site K, (c) misorientation of the surface with respect to the ideal low-index plane.[29]

position in the lattice. At high current densities, adatoms are formed so fast compared with their rate of surface diffusion that many newly deposited atoms cannot diffuse to a favorable lattice site before their motion is impeded by other depositing atoms. The result is a less ordered layer. For most of the electrodeposition process, the metal deposition reaction occurs effectively at an electrode of the electrodeposited metal. Although the rate of this process may be determined using relatively simple current-overpotenial expressions, this does not necessarily mean that the electrodeposition process is simple. The deposition may involve predissociation steps for complexed ions; certain ions in solution may catalyze electrodeposition; and finally, anomalous codepositions can occur in alloy depositions, as mentioned above.

5.3.2 The Effect of Current Density

A schematic representation of current density versus cathode potential is shown in Fig. 5.4.

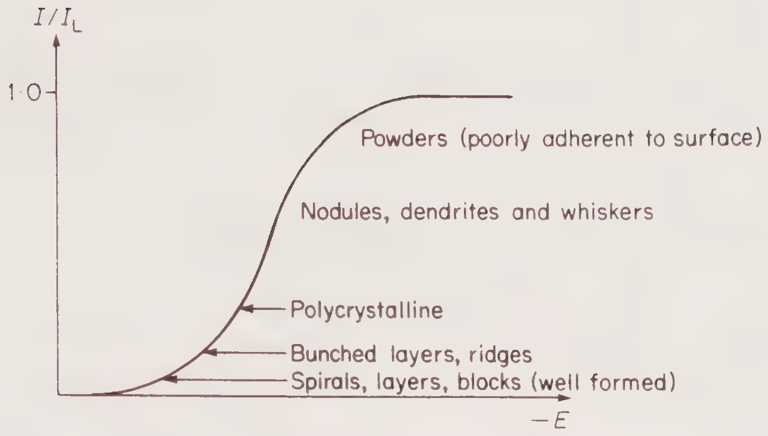


Fig. 5.4. Generalized plot of normalized current display (I/I_L) as a function of current density. Also shown are electrodeposit characteristics that may be encountered at different current densities.[31] I_L is the diffusion limiting current.

As indicated earlier, at low current densities the surface diffusion is fast compared with electron transfer, and the crystal structure and defects, e.g., screw dislocations, are well formed. As current density is increased, with an accompanying increase in overpotential, the formation of additional nuclei occurs at a faster rate. Consequently, the electrodeposit structure will be less ordered, and various large-sized features, nodules, ridges, and steps become more likely.

With further increases in current density, mass transport becomes a predominant factor in the deposition process. The formation of outward growths such as nodules and dendrites is favored under conditions of limited mass transport. Once dendritic growth commences, it tends to dominate the deposition process because of enhanced rates of mass transport to the tips of the dendritic whiskers as a consequence of non-planar, or spherical-type diffusion.

Ultimately, when the limiting current I_{lim} is reached of the deposition process is reached (i.e., there is total diffusion control of the metal ion transport to the cathode in solution), the deposit forms as a powder. The metal ions cannot be supplied fast enough at the I_{lim} limiting diffusion condition. Only part of the current supplied goes to metal discharge while the rest goes to hydrogen evolution. The deposit becomes gray or dark, and powdery. Owing to a rapid increase in pH at the surface as result of water decomposition according to equation (5.8), the metal may deposit as an oxide or hydroxide.

Referring to Fig. (5.4), for compact deposits it is generally desirable to avoid using current densities exceeding about 60% of the limiting current density.

5.3.3 Aspects of the Structure of Electrodeposits

The previous sections discussed some aspects of the initial stages of electrodeposition and the role of current density. Some comments will be made here on the structure of the deposits beyond the initial stages of deposition.

Many electrodeposits obtained under practical current densities are polycrystalline, with many grain boundaries and imperfections. In general, standard electrochemical measurements provide little information about the growth of the electrodeposit after the initiation stage other than current efficiency. Rather, information about thick deposits comes from such *ex situ* techniques as optical microscopy, scanning electron microscopy (SEM), transmission electron microscopy (TEM), atomic force microscpy (AFM), and X-ray diffraction. Nevertheless, the structure and properties of the deposit are completely determined by electrochemical parameters combined with the properties of the plating bath plus, as discussed last, secondary but important factors such as seed layer type and cleanliness. Therefore, if one wants to be able to control the film properties, and be able to tailor them, one must understand the relation that exists between plating solution composition, plating parameters, nucleation and growth of the deposit, grain size orientation, and metallurgical structure and properties. With this knowledge, electroplating technology becomes a predictable science. Frequent complaints that electroplating technology is at best an art occur because some investigators do not understand the importance of these correlations and try to control the properties of deposits by simple manipulations such as changing the current density. This only very occasionally leads to success, hence plating to them is unreliable.

5.3.3.1 Epitaxial growth and preferred orientation

It is possible to obtain epitaxial electrodeposits — a continuation of the crystal structure of the substrate — provided the lattice mismatch between the seed layer and the plated metal is less than 8-10%, and that the substrate surface is free of oxide and not plastically deformed. The most frequently observed initial type of epitaxy is the formation of discrete epitaxial crystallites. Coalescence of the latter then occurs, although the size of the crystallites obtained prior to coalescence tends to be dependent on the composition of the plating solution.

Not surprisingly, epitaxy is likely to be limited to thin deposits, and may be dependent on the plating conditions. Low current density favors epitaxial growth and columnar structure, which is strong in the perpendicular direction and weak in the horizontal direction, since the intergranular regions are the regions of weak cohesion. Therefore, to have uniform strength in all directions, it is preferable to deposit metal with an equiaxial structure, and to avoid columnar structure. The equaxial deposits tend to be ductile, but low in hardness and strength. The columnar deposits show little ductility in the plane due to the separation between the columns, but are usually hard and show very high compressive strength normal to the plane (normal to the columns).

5.3.3.2 Preferred orientation of grains

In electroplating, the grains usually grow perpendicular to the surface; growth parallel to the surface can only occur when it is not restricted by other grains. Deposits may have all three crystallographic axes randomly oriented with respect to the substrate. However, in some cases, grains grow faster in one direction, i.e., they exhibit preferentially-oriented crystal faces. Thus, in some electrodeposits a particular crystal plane is preferentially oriented parallel to their surface. Figure 5.5 shows a sample of a Zn deposit at different thicknesses as it develops preferred orientation as studied by electron diffraction and electron micrographs. In this example, one axis in the crystallites is fixed relative to the surface, the other two being randomly oriented.

It has been pointed out by Rahskov et al.[32] that the overpotential determines which crystal plane exhibits the lowest work of formation and therefore could be preferentially nucleated. Thus the work of formation was lowest for the (111) plane of a face-centered cubic metal at low overpotentials, while other planes were favored at higher overpotentials. At high values of the latter, the work of formation was observed to be similar for all planes, and a randomly oriented structure should be obtained.

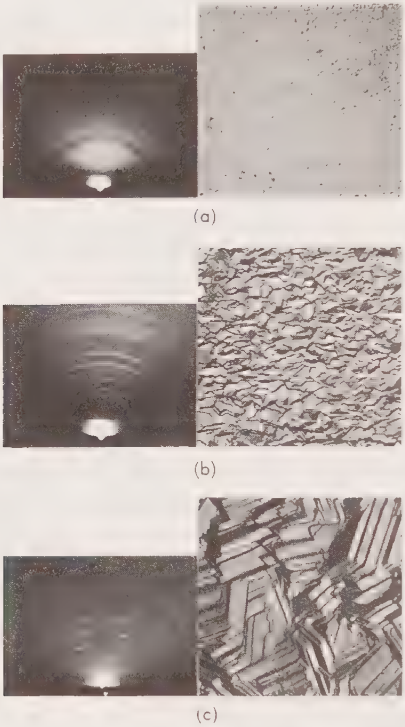


Fig. 5.5. Electron diffraction patterns and electron micrographs showing development of preferred orientation in the electrocrystallization of zinc at 40 mA/cm^2. (a) 700 Å deposit (10,000X), random orientation, (b) 6,300 Å deposit (5,000X), random orientation, (c) 54,000 Å deposit (5,000X), preferred orientation (from Bockris and Razumney[29]).

The phenomenon of preferred orientation can be strongly dependent on the type of plating solution used. For example, Cu deposits obtained from sulfate of fluoroborate solutions are oriented with (110) planes parallel to the substrate surface, while Cu deposited in pyrophosphate and high-temperature cyanide baths exhibits (111) and (200) orientations, respectively.[30] In a Watts nickel bath without additives, grains with a (100) direction perpendicular to the surface grow the fastest, as reported by Ragunathan and Weil[33]; the presence of additives in the solution causes a departure from this behavior, however.

5.3.4 Properties of Electrodeposits

A unique characteristic of electrodeposits is that a very wide range of properties can be observed for even a single-element deposit due to the wide range of deposition parameters available, including the effects of additives. Electroplated metals are often stronger and less ductile than similar materials produced by casting from the melt and by wrought methods. This is mainly due to their rather high dislocation density, small grain size, twinning and the inclusion of additive-related foreign substances. It has been shown in the electrodeposition of Co that the maximum tensile strength is more than four times that achievable in annealed, wrought Co. In general, in the case of Co, Cu, and Ni, grain size refinement during electrodeposition was found to be more effective for increasing strength than grain size reduction achieved by a cold rolling process, for example.[30]

Fine-grained deposits tend to exhibit low ductility. Large-grained deposits show higher ductility, as indicated by considerable plastic deformation before fracturing. However, plastic deformation may be confined to discrete narrow regions of the deposit, which eventually fracture. Large local plastic deformation prior to fracture may also be responsible for many large-grained deposits not cracking under high tensile internal stresses.

Electrodeposits are generally harder than similar materials produced by other methods. This characteristic can be ascribed to a variety of reasons, e.g., higher dislocation densities in electrodeposits, small grain size due to additive-inhibited growth, and even high microstresses. Chromium electrodeposits owe their good wear resistance and hardness to their small grain size. In the area of composite plating where particles suspended in the plating solution are incorporated into the electrodeposit, the incorporation of particles of hard materials such as carbides, diamonds, and certain oxides significantly improves their wear characteristics. Figure 5.6 shows the typical ranges of microhardness that are achieved in electrodeposits.

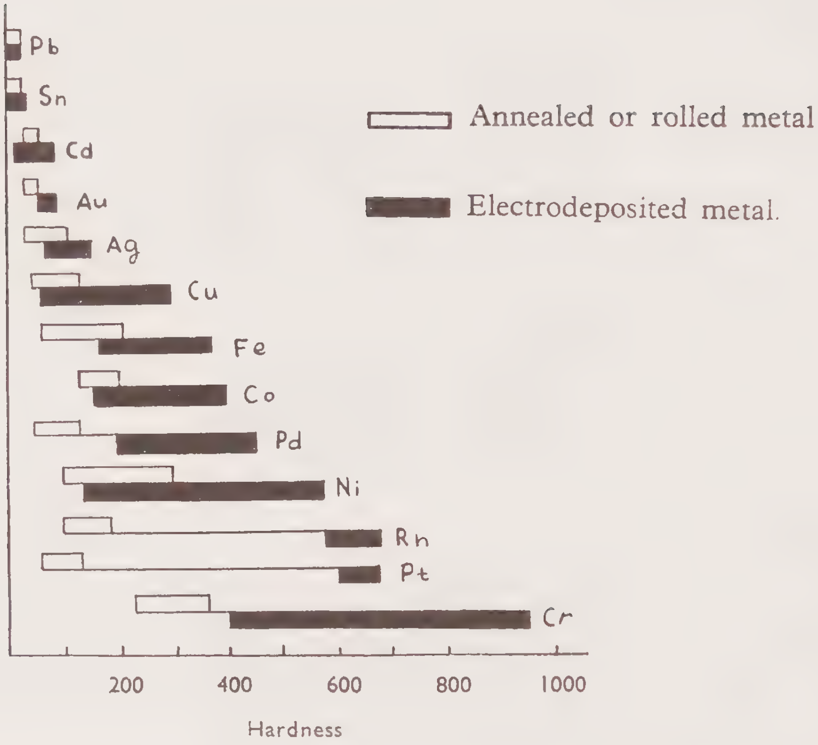


Fig. 5.6. Comparison of hardness data for various metals between the annealed or rolled form and the electrodeposited form.[35]

5.3.5 Stress in Plated Films

By now it should be evident to the reader that electrodeposits exhibit a wide array of unique, interesting, and useful properties, and that there are known techniques to control and even tailor these properties if one is skilled in electrodeposition. One phenomenon that requires some attention, as in many other deposition methods, is internal stress in the deposits. Owing to the ability to control grain size and orientation through tailoring of plating solution composition and control of plating parameters, it is possible to control stresses on plated films over a wide range from compressive to tensile. Figure 5.7 shows some typical ranges for highly-stressed electrodeposits.

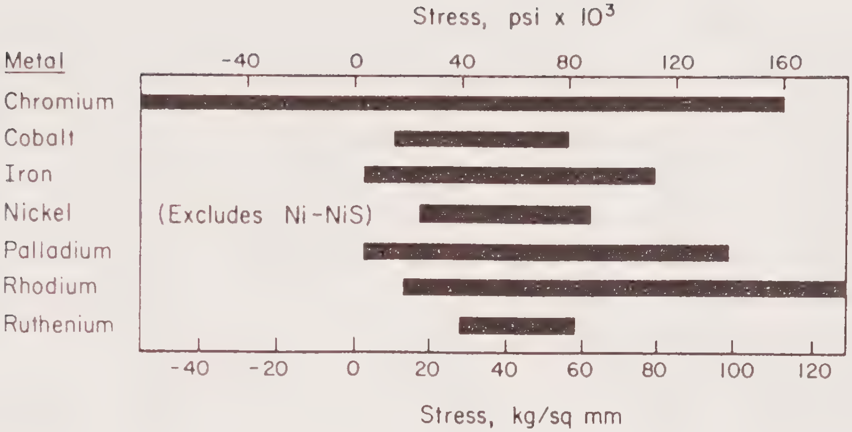


Fig. 5.7. Stress ranges for highly stressed electrodeposited metals.[30]

There are number of types of deposit stress, e.g., extrinsic stresses due to a lattice mismatch between the substrate and the deposit, and intrinsic stresses due to the incorporation of foreign materials such as hydrogen and those derived from additives. The former may be observed in the case of Ni deposits on Cu (high tensile stress; Ni has a smaller interatomic spacing than Cu), and Au deposits on Cu (low compressive stress; Au has a larger interatomic spacing than Cu). When deposit thickness reaches a certain value, extrinsic stresses tend to be relieved through the formation of dislocations.

Tensile or compressive stress may be observed for the same metal from the same plating solution, depending on the plating conditions employed. Metals that exhibit very little stress include Cu, Ag, and Pb; those that show high stress include Cr, Pd, Rh, Fe, and Ni. Generally those metals that deposit in the presence of high hydrogen overvoltage show high tensile stress. However, by using stress-reducing agents, plating conditions can be achieved where low-stress Ni can be obtained, e.g., by using a sulfamate or Watts bath. Metals like Rh or Cr, which are deposited at very high acid levels and high H_2 overvoltage, can be so stressed that they will form microcracks during plating. These cracks usually heal by additional metal plating, which fills the cracks.

The issue of intrinsic stress caused by the codeposition of hydrogen may be understood as follows. Tensile stress can develop due to diffusion of hydrogen out of the deposit surface layer or toward the substrate, causing a shrinkage of the surface or expansion of the layer beneath. Compressive stress is caused by hydrogen remaining in the surface layer, filling voids and expanding them; such behavior may result from the use of certain additives in the plating solution, which change the stress from tensile to compressive. As discussed later, choosing the right additive can overcome problems with hydrogen incorporation, e.g., saccharine in Permalloy solutions. Saccharine can change stress from tensile to compressive in deposits plated from a Watts bath.[34]

5.3.6 Other Properties

A number of other deposit properties are affected by electrodeposition conditions, e.g., density, electrical resistivity, brightness (which was discussed earlier), and corrosion resistance and magnetism (discussed elsewhere). The electrical conductivity of electroplated metals tends to be lower than in cast and annealed samples of the same metal, since defects such as point defects and foreign atoms in solid solution may occur in electrodeposits. Lower density is typically observed in the case of electrodeposits as a result of porosity.

Good corrosion resistance is often a required property of deposits. The classic example of electrodeposits used for corrosion protection is Cr, Ni, and NiFe, all of which are used extensively in the automotive industry. Corrosion protection by electroplating was one of the early applications for electroplating. Tons of electroplated metals are used annually for corrosion protection purposes. There is a high level of understanding of this technology, and the reader is referred to a number of books on this subject.[36-38]

A few phenomena to be avoided in electrodeposits where good corrosion resistance is desirable are the following: (a) porosity, which tends to reduce corrosion resistance has to be avoided — pores may result from adhering gas bubbles that prevent deposition; (b) significant variations in the composition of grains, or even their size, may reduce corrosion resistance; (c) crevices surrounding nodules, and those due to preferential adsorption of additives which slow the plating process can be preferentially attacked in harsh environments, and pits, which are the major villains in corrosion, can readily occur in these areas.

5.3.7 Additives and Plating Formulations

It would be possible to list many different bath formulations[5,30], including the additives if space permitted. Workable formulations can be prepared by those who have an understanding of electroplating technology and solution chemistry. To proceed without such knowledge may lead to frustration and discouragement. It is the opinion of the authors that for those who have no formal training in electrochemistry or electrochemical engineering, it is best to employ plating solutions developed by the various vendors with experience in this field. These vendors have invested a lot of effort in developing solutions for a wide variety of applications. Their knowledge is particularly valuable if the plating bath formulation has been sold for a long time, and considerable experience has been built up in the course of its use for the specific application.

The type and the concentration of additives have to be fine tuned for a specific application. For instance, the plating solutions that were developed for decorative plating or for corrosion protection generally cannot be readily used for fabricating electronic or magnetic devices. However, with proper adjustments of plating conditions and of additives, which are usually derived from certain classes of organic compounds by those who understand the science behind the electrochemical plating solution formulation, it is possible to achieve the desired metal properties. Some solution vendors add several different additives even from the same class of compounds. They thereby occasionally achieve a synergistic effect, which distinguishes their plating solution from others. By doing so, they may improve a particular property of the film that makes it more useful for some specific application.

One disadvantage of using commercial solutions is that the additives they contain are proprietary, which at times may not be readily monitored. However, it is often the case that these additives determine if a plating solution meets the needs of a particular application. A disturbing aspect of vendor plating solutions with additives is that details of their mode of operation are not always known to the user. The larger and more reputable vendors put a considerable amount of effort into understanding the effect of additives, and into learning how to precisely control the plating solution. The vendors are therefore able to provide detailed instruction on the proper use of the solution, the replenishment schedule, and means of testing the solution to ensure proper operation under normal conditions.

However, in the case of some additives it may not be clear to the user which additive or additives create the desired effect, or whether it is the decomposition product of the additive that is responsible. Some vendors of plating solutions will share this information with the users under a confidential agreement so as not to jeopardize their business.

Some comments may be made on the mode of operation of additives in plating solutions. The additives are generally surface-active chemicals. They adsorb strongly at the active sites on the cathode and consequently they or a product of their cleavage are often incorporated in significant concentrations into the deposits at grain boundaries, especially when the plated deposit has a high surface energy, as reflected in a high melting point. Such inclusions of additive in the deposit modify the properties of the deposit, e.g., grain size and structure, preferred grain orientation, mechanical and magnetic properties, hardness, elasticity, and maximum elongation. The incorporation of such foreign material in grain boundaries can lead to obstruction of the movement of dislocations, thus increasing deposit strength and decreasing ductility or increasing hardness. Corrosion resistance may also be affected by the incorporation of additives or their products.

A further consequence of the adsorption behavior of additives on cathode surfaces is that they increase the deposit overpotential and change the slope of the current (logarithmic value) vs. potential plot, the so-called Tafel slope. This is probably due either to the need for electron transfer to occur through the adsorbed layer of additive or its product on the cathode, or to complex formation with the metal ion at the cathode. As discussed earlier, changing the overpotential at the cathode can have a major effect on the rate of nucleation and properties of the electrodeposit.

5.3.8 Brighteners

An important function of certain plating solution additives is to provide bright deposits for decorative purposes in which light is reflected specularly. In such deposits, the microscopic roughness of the deposit is low compared with the wavelength of the incident light so that it is reflected rather than scattered. Generally, protrusions and crevices on the electrodeposit surface should not vary in height more than the wavelength of light. Large grained deposits tend not to be bright because of their boundaries, and the height of features on their surfaces often exceeds the wavelength of light. However, although large grains may be semibright if they consist of layers that are largely parallel to the substrate surface, small grains free from nodules or depressions are a necessity for brightness.

Since brightening agents are used in relatively high concentrations (e.g. 10^{-2} M), In bringing about the formation of a fine-grained deposit, brighteners apparently act by modifying the kinetics of nucleation. Examples of additives often used are aromatic sulfonates in Watts-type Ni solutions (weakly acidic chloride or sulfate baths), and additives such as coumarin discussed by Safranek.[30]

5.3.9 Levelers

Another important role of additives is that of leveling, which is the production of a macroscopically level deposit on an uneven substrate surface. This type of additive acts by adsorption onto regions where otherwise there would be a rapid rate of electrodeposition i.e., small regions of the surface of a size comparable to the thickness of the diffusion layer for the levelling additive, which protrude into the solution. Preferential adsorption also occurs at dislocations because of a higher free energy of adsorption. The adsorbed additives act to slow the kinetics of electron transfer. Figure 5.8 shows the leveling that can be achieved when electroplating nickel over a specially grooved surface.

5.3.10 Stress-Reducing Agents

Additives may influence structural properties such as stress, and also may be used to improve wetting. Additives can play an important role in decreasing stress in the deposit, whether due to lattice mismatch with the substrate, or to hydrogen incorporation in the deposit, or sulfur incorporation. Hydrogen, which in most cases is evolved in a reaction parallel to electrodeposition, can be incorporated into the deposit in significant concentrations and can be a source of unwanted stress. Nickel is capable of absorbing up to 40 vol % of hydrogen. Since nickel is electroplated at a considerable hydrogen overvoltage, this means that during plating (when current is flowing), the Ni deposit is loaded with hydrogen. It may be under an equivalent of several atmospheres of H_2 pressure during the electroplating process. When the current flow is interrupted, the H diffuses to the surface of the deposit where it combines to form H_2 mole-

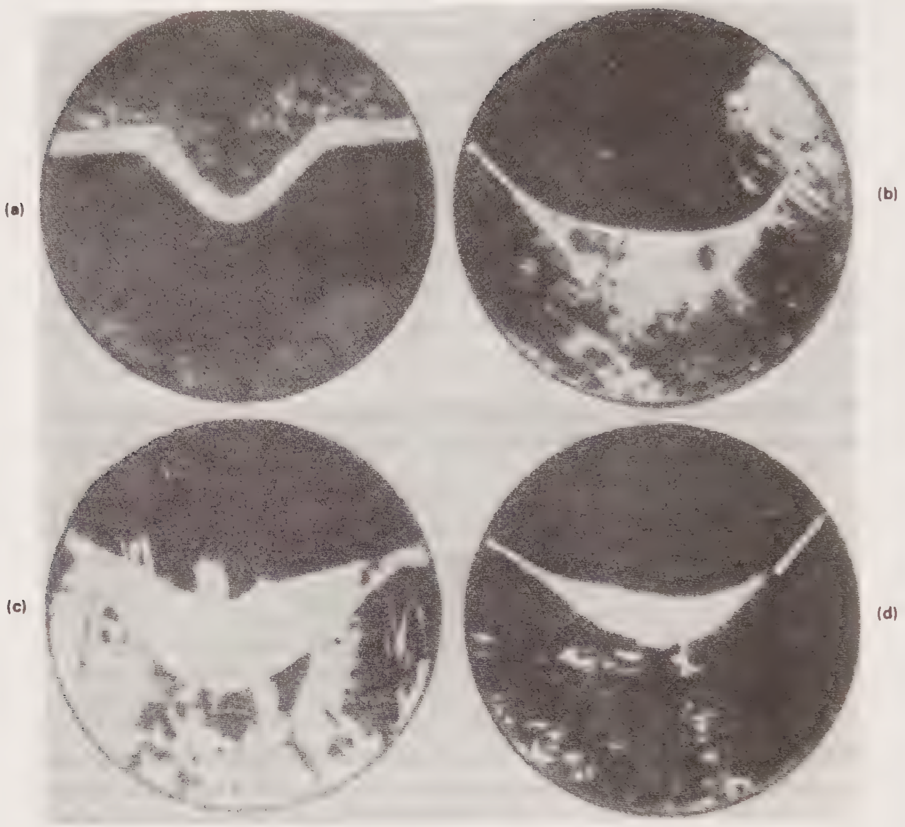


Fig. 5.8. Photomicrographs of nickel deposits produced from solutions containing no organic additives. (a) 2 g/l of paratoluenesulfamide + 0.05 g/l of coumarin, (b) + 0.1 g/l of coumarin, (c) + 2 g/l of coumarin, (d) Mag. = 315X.[39]

cules and escapes as H_2 gas. Following removal of the hydrogen from the Ni, the latter is under tensile stress. This can be overcome by using sodium saccharine as an additive. Indeed if enough sodium saccharine is used, it is possible to deposit nearly stress-free nickel or to change the sign of the stress.[33]

In other instances the use of wetting agents promotes the release of hydrogen gas bubbles from the cathode surface when they are very small. Pits, which result from clinging H_2 bubbles are also eliminated in this manner. The incorporation of hydrogen is more likely in plating solutions in which the hydrogen overpotential is very high because the metal deposits at a high negative potential or it is so strongly complexed that the electrodeposition occurs at rather negative potentials.

5.4 CURRENT DENSITY UNIFORMITY

In electrodeposition, deposit thickness and alloy composition are determined by local conditions of current density and mass transport. When current density varies across a substrate it may cause thickness nonuniformities on the scales of (a) the substrate, (b) the pattern, or (c) the individual feature. When electrodeposition is followed by planarization, many of the issues of uniformity are not of major concern for single-element deposition. The situation is more complex when depositing an alloy, since alloy composition may be a strong function of current density. This is particularly true for alloys of Fe, Co, and Zn with Ni. The magnetic properties are strongly dependent on alloy composition and other current-density dependent properties such as grain size, structure, and crystal orientation. In plating through resist masks, if there is no planarization step in subsequent processing, there may be areas where high current density results in plating beyond the surface of the resist, making removal of the seed layer difficult in these localities.

5.4.1 The Various Types of Current Density Distribution

It was indicated earlier that it is important to use a current density within a certain range for best deposit properties. However, how the current is distributed over the surface is important since this determines the *local current density*. The latter determines local deposit thickness, which may be different between peaks and valleys, and small and large features. While much practical knowledge has been built up over the years on how to handle this issue, mathematical modeling has begun to be applied to gain a better understanding of the factors influencing current distribution on surfaces with a range of features, and to be able to predict and optimize thickness uniformity. Aspects of the present understanding of current distribution are summarized here.

Prior to discussing current density uniformity in electrodeposition, the different types of current distribution, namely, primary, secondary and tertiary, are described. These terms are central to any discussion of deposit uniformity, since they appear in most of the technical literature dealing with uniformity in electrodeposition. A number of literature sources may be consulted regarding these; however, the chapter by Pinkerton gives a practical introduction to these quantities, including potential theory, which helps one to understand current distribution.[40]

When an electric potential (i.e., voltage) is applied between two electrodes in an electrolyte, each point in the electrolyte assumes some potential intermediate between that of the two electrodes. Since the conductivity of metal electrodes is several orders of magnitude higher than that of the electrolyte, it is assumed to a first approximation that every point on the surface is at the same potential. Thus the surface of a metal electrode is considered to be an *equipotential surface*. Equipotential surfaces will also be found in the electrolyte surrounding each electrode. As one moves away from the electrode surface, the shape of these equipotential surfaces changes.

It is a fundamental tenet of potential field theory as applied to electrode surfaces that current enters or leaves the electrode at any point in a direction perpendicular to the surface at that point. Reference will be made below to current lines, which are imaginary lines that cross equipotential lines (2D representation of equipotential planes) in a perpendicular manner, whether in solution or at the electrode surface. The ability to qualitatively map the potential field, or better still to mathematically model it, around a relatively complicated electrode structure is important in optimizing cell geometry.

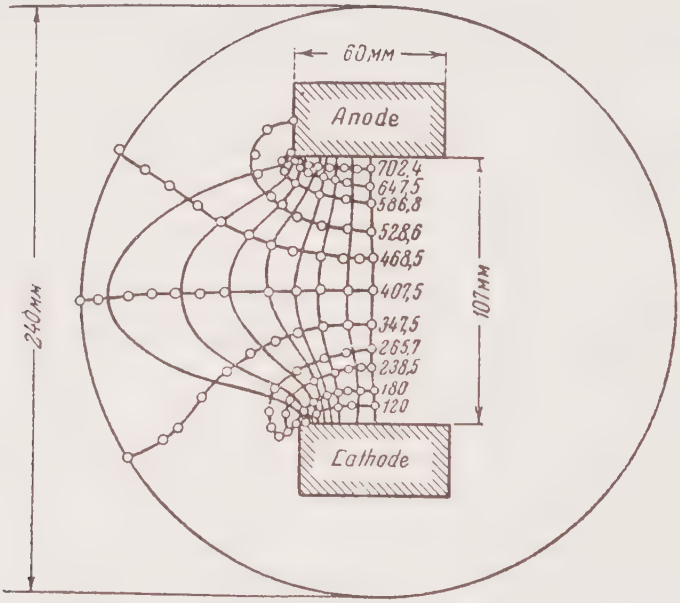


Fig. 5.9. Distribution of equipotential lines and lines of force in an electrolyte of composition 125 g/1 $CuSO_4.5H_20$, 4•16 g/1 H_2SO_4.[35]

Figure 5.9 shows the equipotential lines and lines of force (i.e., current lines) in a simple electroplating cell containing a fairly concentrated acid Cu plating solution for an applied current density of 20 mA/cm^2. As can be seen in this figure, the equipotential and current lines are not simply confined to the volume of the solution between the cathode and anode, but rather "bulge" out into solution. In this simple example, deposit thickness will be greatest at the edges of the cathode where the equipotential lines are most concentrated. Thus, as discussed later, the cell, the cathode, and any resist pattern on it have to be designed in such a manner as to give an acceptable uniformity of deposit thickness. As will be seen, it is usually possible to accomplish this.

5.4.1.1 Primary current distribution

The primary current distribution is the current distribution obtained across an electrode in the absence of polarization, i.e., the extent that the electrode potential has departed from the standard potential for the reaction. This distribution is determined completely by the geometry of the plating cell, and is independent of the properties of the plating solution provided the properties are uniform throughout the region encompassing the plating electrode.

$$\frac{I_A}{I_B} = \frac{d_B}{d_A} \tag{5.10}$$

where I_A and I_B are the current densities, respectively, at distances d_A, d_B from the anode measured along *the current lines* for these two points. When an electric potential (i.e., voltage) is applied between two electrodes in an electrolyte, each point in the electrolyte assumes some potential intermediate between that of the two electrodes. Since the conductivity of metal electrodes normally exceeds that of the electrolyte by many orders of magnitude, it is usually assumed to a first approximation that every point on the surface is at the same potential. Thus the surface of a metal electrode is considered to be an equipotential surface.

Equipotential surfaces will also be found in the electrolyte surrounding each electrode. As one moves away from the surface, the shape of these surfaces changes. It is a fundamental tenet of potential field theory as applied to electrode surfaces that current enters or leaves the electrode at any point in a direction perpendicular to the surface at that point.

5.4.1.2 Secondary current distribution

When one or more phenomena occur to cause polarization, i.e., an increase in potential for a fixed current density, at the electrode surface, a local or secondary current distribution results at the surface that is different than the primary current distribution. These phenomena, which cause an increase in the resistance to the flow of current, may be, for example, a change in the cathode film if present, a reduction in the concentration of the conducting ions in the electrolyte, a change in the concentration of metal ions in the solution near the electrode, or gas bubble shielding.

A simplified expression for the secondary current distribution is obtained by adding the term kdE/dI as a second term to both the numerator and denominator of the right-hand side of the equation given above for the primary current distribution. Here k is the conductivity of the solution, and dE/dI is the slope of the cathode polarization curve, i.e. the potential vs. current density curve, between the two current densities. This term has the dimension of length. The combined effect of polarization and solution conductivity is therefore equivalent to adding an equal solution length to the current path lengths d_A and d_B. Note that concentration polarization can alter the conductivity of the solution adjacent to the cathode.

In general, the secondary current distribution acts to oppose the tendency of the primary current distribution to concentrate current flow on high current density areas. Thus the secondary current distribution ratio is generally more uniform than the primary current distribution. In addition, in the secondary current distribution region, the leveling additives tend to adsorb on the high current density areas. Thus the films produced in this region tend to have a more uniform thickness, and to show a considerable amount of leveling.

5.4.1.3 Tertiary current distribution

In the tertiary current distribution region, the current is controlled by the rate of supply of the metal ions to the surface. If the concentration of the metal ions in the bath is designated as C_b and the concentration of the ions at the surface is C_s, the limiting current can then be expressed as

$$I_{\text{lim}} = \frac{nFD(C_b - C_s)}{\delta} \tag{5.11}$$

where n = the charge of the metal cation, F = the Faraday constant, D = the diffusion coefficient (cm^2/s), and δ = the diffusion layer thickness in centimeters. Under normal plating conditions in the secondary current distribution region, $C_b > C_s > C_o$, where $C_o \approx 0$ is the limiting current concentration at the surface. Since the limiting current depends on the bulk concentration and diffusion coefficient, the charge of the cations and the Faraday constant, all of which are constant in a plating solution the only variable that one has is the diffusion layer thickness.

5.4.2 Diffusion Layer

A few comments on the diffusion, or mass transport or depletion, layer will be made now. In the case of electroplating, this is the region of the solution close to the cathode where the concentration of the metal ion being electrodeposited is less than the value in the bulk of the solution, C_b. The thickness of this region may reach tens of microns in the absence of convection in the solution. Convection always decreases the thickness of the layer by increasing the transport of metal ion to the cathode, and is virtually always employed in practical electroplating. The bulk concentration of the metal ion in bulk solution is only maintained at the cathode in a few special cases, for example, in the case of very low current densities and if the value of C_b is very large. As the diffusion layer thickness is decreased by convective agitation, solution flow or other means such as ultrasonic agitation, I_{lim} can be changed by as much as one to two orders of magnitude.

In practical electroplating, some depletion of the metal ion occurs at the cathode as shown in Fig 5.10. In this schematic representation of the region close to a cathode, significant depletion of the metal ion has occurred as a result of electrodeposition; it is evident in this case that the concentration of the metal ion at the surface, C_s, is about 50% that of C_b. The diffusion layer thickness is usually characterized by its *effective* thickness, δ_H, which is measured from the electrode to the point of intersection of the two tangents (dashed lines) shown in Figure 5.10. The rate of diffusion of metal ion in *mol/cm²/s*, is proportional to the concentration gradient at the electrode, and is given by $D(C_b - C_s)/\delta_H$, where D is the diffusion coefficient of the metal ion.

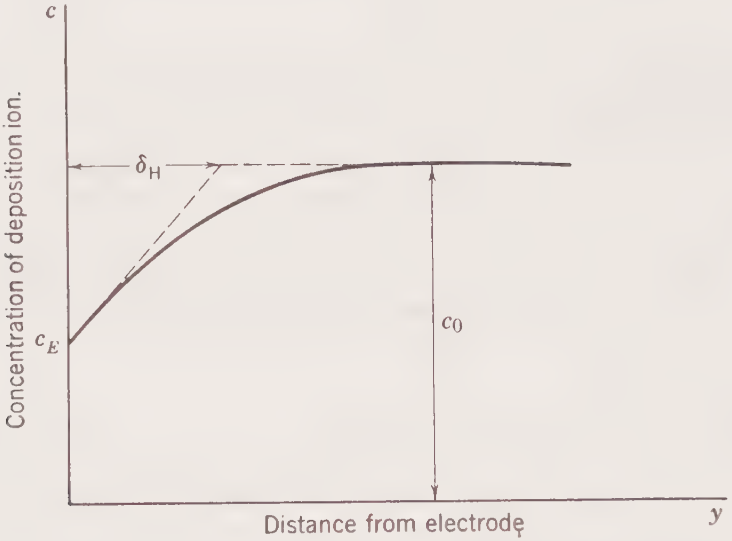


Fig. 5.10. Schematic representation of metal ion concentration near a cathode during deposition: C_0, concentration of ion in bulk solution; C_E, concentration at the electrode surface; δ_H, effective thickness of diffusion layer.[5]

The diffusion gradient, or extent of depletion of metal ion, at the cathode will tend to increase with increasing current density of electrodeposition. In extreme cases where all the metal ions are being reduced as fast as they reach the cathode, the value of C_S will be zero. As will be seen later, this is not a desirable situation in electroplating since it leads to "burnt," or powdery deposits.

It was stated earlier that in practical situations it is not desirable to plate at current densities higher than 60% of the limiting current density because of the deleterious effect of such current densities on the structure of the deposit. From the discussion in this section, it is clear that if one is prepared to increase the rate of agitation and thereby decrease the diffusion layer thickness, the limiting current I_{lim} can be increased and good-quality deposits can be obtained at very high current densities. The other alternative is to increase the concentration of the metal ions in solution. The issues of the diffusion layer thickness, the diffusion constant, and the bulk concentration of metal ions become very important when plating through resist masks, and in particular in the LIGA process in which one may have to plate into recesses with up to 500:1 height-to-width aspect ratios, where it is impossible to agitate all the way to the bottom of the trench.

5.4.3 Uniformity of Current Density Across a Substrate

The boundary between a conductor to which current is applied and an insulator causes a nonuniformity in the potential field. The consequent nonuniformity in plating current is evidenced by nonuniformity in the deposit thickness. Redistribution of current from a uniform planar distribution is often called "current crowding." On the substrate scale, the current crowding at the substrate edge can cause substantially higher thickness there. The difference in deposit thickness across a substrate may be as much as 50%, and can extend a significant distance from the edge. In irregularly shaped cathodes, this can have a pronounced effect as seen in Fig. 5.11.

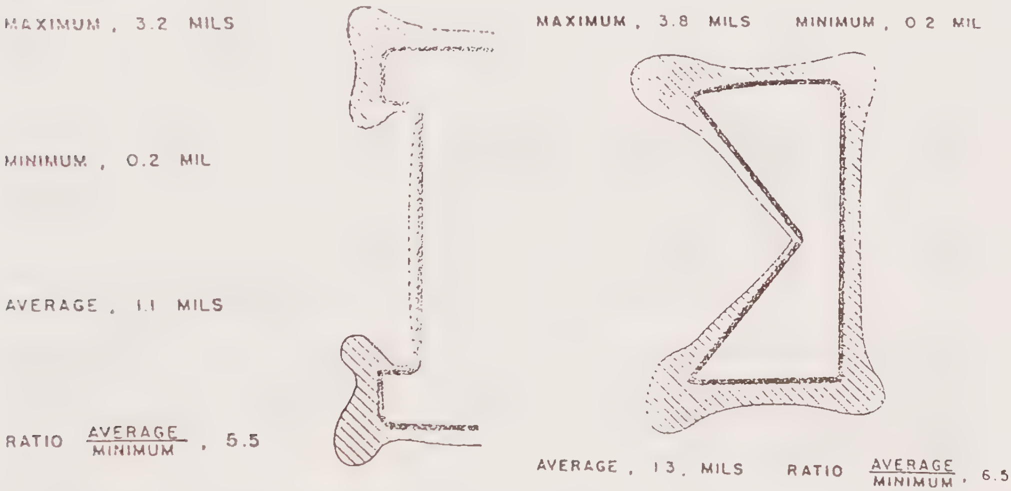


Fig. 5.11. Nonuniformity in Ni electrodeposit thicknesses on two different irregularly shaped cathodes.[6]

These effects can be minimized by a number of methods, such as *current thiefs* and local *current shields.*

In the case of planar substrates, independently-powered auxiliary electrodes, often known as current deflectors or thieves, that surround and are coplanar with the cathode, are employed to diminish this edge effect, and to achieve deposits with much improved, across-the-wafer uniformity. This problem has been studied by a number of groups. It has been modeled extensively by Mehdizadeh et al.,[42] who studied the situation that applies in a paddle cell.[43,44] In its simplest form, this is a rectangular cell containing a reciprocating paddle that sweeps to and fro at a controlled distance of a few millimeters above the horizontal cathode surface at speeds up to 1 Hz. The rectangular anode is placed parallel to and above the cathode and auxiliary electrode, and the reciprocating paddle. The walls of the cell are adjacent to the outer edges of the auxiliary and anode electrodes, and hence they ensure that current lines are contained within the rectangular region encompassing the electrodes.

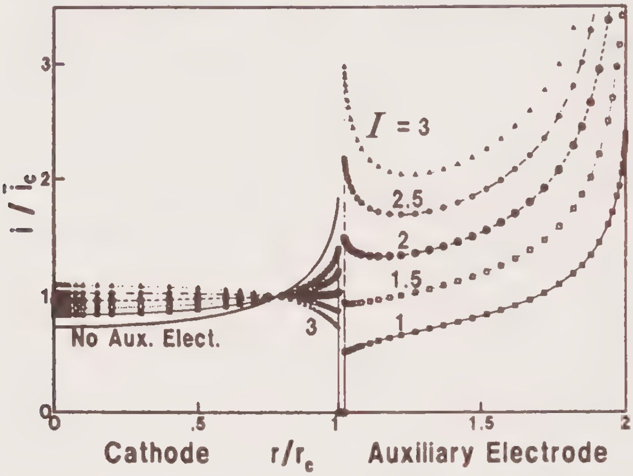


Fig. 5.12. Normalized secondary current distributions at the cathode and auxiliary electrode with Tafel-type polarization for different I values, the ratio of auxiliary electrode current to cathode current density.[42] i_C = average current density on cathode; r = radial position on cathode; r_C = radius of cathode.

Figure 5.12, taken from Ref. 42, shows the effect of auxiliary electrode current on the secondary current distribution at the cathode in a paddle cell.[43,44] When no current is applied to the auxiliary electrode, there is major current density nonuniformity at the cathode that extends a considerable distance from its edge. Application of increasing currents to the auxiliary electrode (shown in Fig. 5.12) as *I*, the ratio of auxiliary current to cathode current) increases the cathode current density uniformity until an essentially uniform current density is obtained for *I* = 2.5. Further increases in *I* lower the current density near the edge of the cathode. An optimum value of *I* needs to be established for each cathode/auxiliary electrode case.

5.4.4 Local Current Distribution Uniformity

In electrodeposition on photoresist-patterned substrates, the resist pattern may strongly influence the uniformity of the deposit thickness. The pattern (i.e., a repeating unit) scale, the current distribution caused by the spacing of the resist pattern or groups of patterns, and the feature scale, the current distribution within a pattern, need to be considered in assessing the thickness distribution when plating through resist. Practical experience and modeling have demonstrated that significant local non-uniformity in deposit thickness can occur[1]: however, much progress has been made in improving the uniformity of local and macroscopic deposit thickness.

5.4.4.1 Resist pattern-scale uniformity

The density of the active plating area may vary among patterns, resulting in pattern-to-pattern variations in deposit thickness. In addition, the patterns are separated by insulating areas where no plating occurs. The resulting higher current density at the pattern edge causes a nonuniform across-device thickness and composition.

The variation in thickness between groups of features in electrodeposition has been studied by Mehdizadeh et al.[45,46]. Using potential theory, these authors modeled the behavior expected for electrodeposition with a secondary current distribution, where there is little mass transport control. They verified their modeling results experimentally using resist-patterned substrates containing 10-, 35-, and 140-μm-wide electrochemically active lines formed in Cr/Cu seed layers.[46] Electrodeposition was carried out in a paddle cell. Good agreement was obtained between the model prediction and the electrodeposit thicknesses. The results indicate that when mass transfer control is not significant, a patterned area can be characterized completely by its overall active electrochemical area and modeled effectively.

These authors[46] also found that the local current density is generally higher where there is a low plating area density, i.e., where much of the seed layer is covered by resist. This variation in deposit thickness uniformity between regions of varying active plating areas, or photoresist density, may be exacerbated by low Wagner numbers[x]. Thus, among other things, the choice of an appropriate plating solution (e.g., one of high conductivity) may help to improve uniformity. Careful design of the exposure mask, perhaps with the inclusion of dummy plating areas could be used to overcome this type of nonuniformity.

A tertiary current distribution case, or where there is significant mass transfer control of deposition, may apply in certain electrodeposition solutions. Mehdizadeh et al.[46] extended their model to include the effects of mass transfer control and tested it experimentally. Whereas reasonable agreement was obtained between theory and experiment for 10 μm wide lines, the agreement was less at line widths much greater than the diffusion layer thickness, possibly because of the enhancement effects of transverse convection.

The modeling results illustrate that an understanding of deposit thickness uniformity can be obtained, but it should be kept in mind that the relatively simple case of an additive-free Cu solution was treated.[46]. It is difficult to model features much larger than the diffusion layer thickness, probably due to intrafeature, lateral convection.[46] It is even more difficult to model plating in baths with one or more additives, or in baths involving alloy deposition, especially for materials like Permalloy, for which anomalous deposition rates for one component are observed.[8]. Finally, the resist employed in HI-MEMS is usually an order of magnitude thicker than those employed by Mehdizadeh et al..[45,46] The alloy composition varies between larger and smaller pattern features, and may also vary between the bottom of deep features and the top of the deposit because of the variation of the mass transport conditions between the top and the bottom.

5.4.4.2 Feature-scale uniformity

A discrete resist pattern usually consists of a large number of features, eg., conductor lines and vias. The features that make up an electrochemically active pattern are not all the same size, and are not uniformly spaced. Uniformity of deposition is governed both by the size of the individual features and by their location. Feature-scale uniformity has been discussed at some length by Romankiw.[1]

---

[x] A *Wagner number* is a dimensionless group that represents the ratio of polarization resistance to the solution resistance. The larger it is, the more uniform is the current distribution, even in the presence of significant nonuniform geometry. A Wagner number of 0 corresponds to primary current distribution, i.e., it is the transfer of changes through the solution and not through the interface that determines deposition uniformity.

Thickness nonuniformity for electrodeposits is greatest for isolated features because of "current crowding." Large features may be thicker at their edges, especially if they at a distance from other active features. Small isolated lines will be thicker than more closely spaced small lines. The situation is worsened in the tertiary current distribution region, since radial, or nonplanar diffusion, amplifies the nonuniformity.

5.5 ELECTROPLATING IN THE ELECTRONICS INDUSTRY

5.5.1 Significance of Plating-Through-Mask Technology

Up to now we have discussed electrodeposition on large surfaces in continuous film (sheet form), with an occasional reference to either the plating-through-mask or LIGA technologies. It is the process of electroforming through thick resist masks that makes possible the unusually tall 3D structures in LIGA.

The invention and development of through-mask-electroplating technology has had a profound influence on the direction in which electronics has evolved. It made extension of thin film metallization technology into the third dimension possible, simple, and cost effective. Devices that in the late 1960s were only a dream or the products of wishful thinking because of the requirements of high component and conductor density, became reality earlier than expected due to the invention and development of the electroplating-through-mask technology. Chemical etching and dry processes, because of the nature of the mechanism of the pattern transfer, to this day are unable to match the fidelity of the pattern transfer that is achieved by electroplating through masks. Electroplating through masks virtually replicates the lithographically generated polymer masks (molds) down to molecular dimensions. It is the only pattern transfer process that has kept pace with every development in lithography: shrinking lateral dimensions down to 0.03 μm; increasing height-to-width aspect ratio to as high as 500:1; and maintaining vertical walls and parallelism of two adjacent walls at a separation of only 1000 Å through thicknesses of tens of microns. None of these is possible using other pattern transfer techniques.

The success of electroplating through resist masks has occurred not only because the process is relatively simple, cost effective, and easy to set up and carry out, but because a considerable amount of work has been put into its understanding and development of it between 1960 and today: much work was done to understand the reaction mechanisms, bath chemistry and process control, and considerable effort was put into the invention, engineering, design, and fabrication of tools which could meet the demands of the electronic devices. More recently, a mathematical model was developed which made through-mask plating more *predictable*.[42,45,46] The understanding of plating through masks and of process control came as a result of a conscious effort and planning at the IBM T.J. Watson Research Center to develop this area. This was done through a concentrated effort on both the fundamental and engineering aspects of it at the T.J. Watson Research Center, through lectures at the leading Universities to point out the challenges to faculty and graduate students, and through publications.[47-49] This effort helped turn what had been considered the art of plating into science.

The basic breakthrough came when it was realized that it is possible to accurately replicate the resist pattern, or the mold, when only the bottom of a mold is metallized with thin conductor, called the seed layer or plating base, and when the walls of the resist mask are nonconducting. It was possible to fill even very small vias completely with the plating solution, and to effect electrodeposition by passing a suitable negative, or cathodic, current through the thin seed layer. This produced an exact replica of the mold. It was also discovered that wetting of the hydrophilic walls of the Novolac family of resists and of PMMA by typical plating solutions was readily achieved. Frequently the capillary forces assisted the process.

It was demonstrated that the electroplating reproduces extremely faithfully the resist mold and promised extendibility into the nano scale for both electronic and MEMS devices. We coined the name for this process "a room temperature metal injection molding process".

5.5.2 Impact of Plating-Through-Mask Technology

Plating-through-mask technology made it possible to build the first thin film heads in 1969-70[50]; to bring them into manufacturing and announce a product in 1979[51]; made possible high resolution bubble memory in the 1970-75 time period[52 - 54]; enabled fabrication of the first useable electroplated gold X-ray lithography masks with sharp edge definition in 1974[55,56]; and made possible fabrication of the first 60:1 height-to-width aspect ratio electroplated gold structures separated by only 1000Å space, using PMMA resist which was exposed using the DESY synchrotron[56]. All these accomplishments were precursors of what is known today as plating-through-mask technology in electronics, and as LIGA in MEMS. Figure 5.13 shows some of the key dates and developments which occurred since the first introduction at IBM T.J. Watson Research Center of the *through-mask-plating* technology in the thin film heads and x-ray lithography. Some of the early publications and recent reviews of the development of the plating-through-mask technology are in Refs. 1,2,42,47-51.

The remainder of this chapter will concentrate on citing some selected examples from the literature and discussion of various aspects of the use of this technology in electronics and in MEMS. It will also discuss the challenges and issues that one encounters in using electroplating in electronics and in MEMS when plating into very deep, or high-aspect-ratio circular holes and trenches.

5.5.3 Plating Through Lithographic Masks — Process Details

The typical process steps used in fabrication of structures using the plating through mask technology are demonstrated in Fig. 5.14. The substrate is usually a dielectric that is typically metallized first with an adhesion layer of 50 to 500Å of refractory metal (such as Cr, Ti, Ta) followed immediately in the same vacuum pump down, by a 400 to 5000Å layer of Ni, Au, Cu or NiFe as a cathode plating surface, or seed layer. The conductive metal thickness depends on the size of the wafer and the desired ultimate thickness of the pattern. In the case of Ni or NiFe, occasionally the adhesion layer is omitted, but then the Ni or NiFe is deposited at 200 - 300°C to form a nickel oxide and/or iron-oxide bond between the metal and the dielectric.

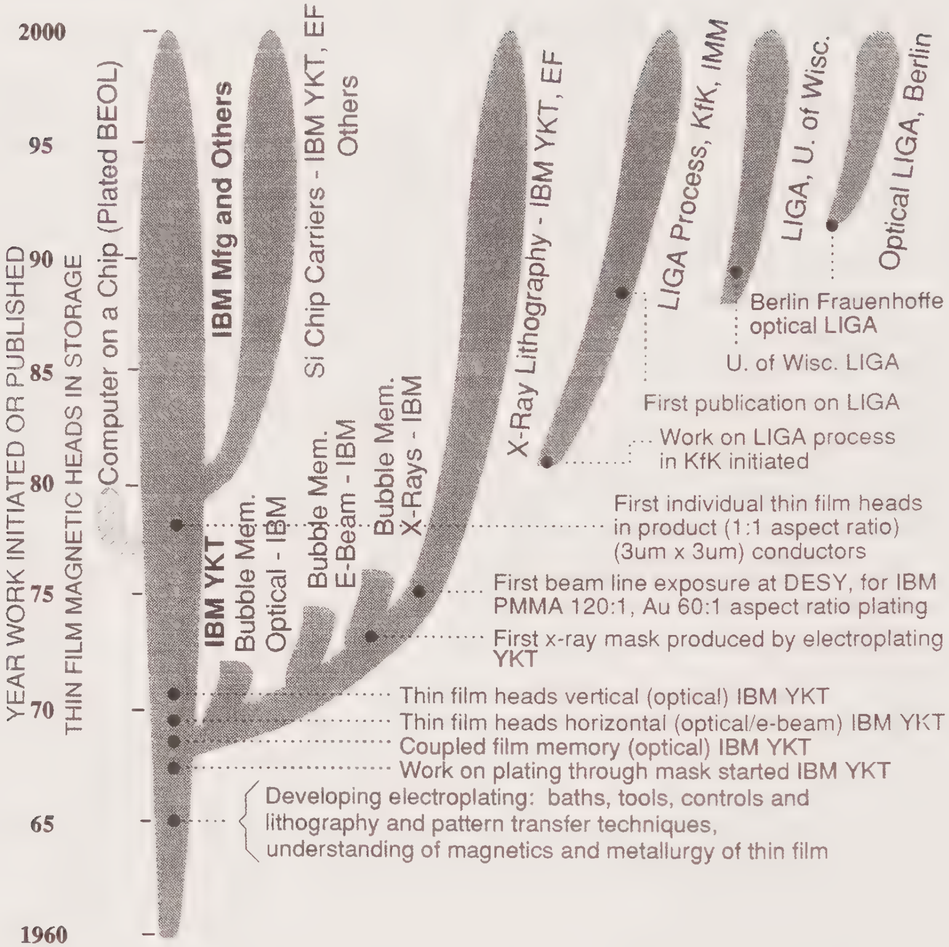

2000
95
90
85
80
75
70
65
1960
YEAR WORK INITIATED OR PUBLISHED
THIN FILM MAGNETIC HEADS IN STORAGE
Computer on a Chip (Plated BEOL)
IBM Mfg and Others
Si Chip Carriers - IBM YKT, EF Others
IBM YKT Bubble Mem. Optical - IBM
Bubble Mem. E-Beam - IBM
Bubble Mem. X-Rays - IBM
X-Ray Lithography - IBM YKT, EF
LIGA Process, KfK, IMM
LIGA, U. of Wisc.
Optical LIGA, Berlin
Berlin Frauenhoffe optical LIGA
U. of Wisc. LIGA
First publication on LIGA
Work on LIGA process in KfK initiated
First individual thin film heads in product (1:1 aspect ratio) (3um x 3um) conductors
First beam line exposure at DESY, for IBM PMMA 120:1, Au 60:1 aspect ratio plating
First x-ray mask produced by electroplating YKT
Thin film heads vertical (optical) IBM YKT
Thin film heads horizontal (optical/e-beam) IBM YKT
Coupled film memory (optical) IBM YKT
Work on plating through mask started IBM YKT
Developing electroplating: baths, tools, controls and lithography and pattern transfer techniques, understanding of magnetics and metallurgy of thin film

Fig. 5.13. Genesis tree of the electroplating-through-mask technology.

The seed layer is coated with an organic polymer, an optical, e-beam, or x-ray resist, which on being exposed through a mask or written by an e-beam, leaves certain areas depolymerized and ready to be dissolved away by the developing solution. After development, thorough water rinsing and drying, any remaining residue of the resist may be removed by ashing with a plasma containing oxygen or forming gas. The open areas are then electroplated.

If a Novolac-type positive working resist is used, the plating solution usually readily wets and fills even a very high-resolution, tall patterns. Upon completion of the plating, the remaining resist is removed using a second exposure and development, or by other suitable means, e.g. using a suitable solvent, or a plasma process. It is worth noting that if a Novolak type resist is used it is possible to do a second alignment, second

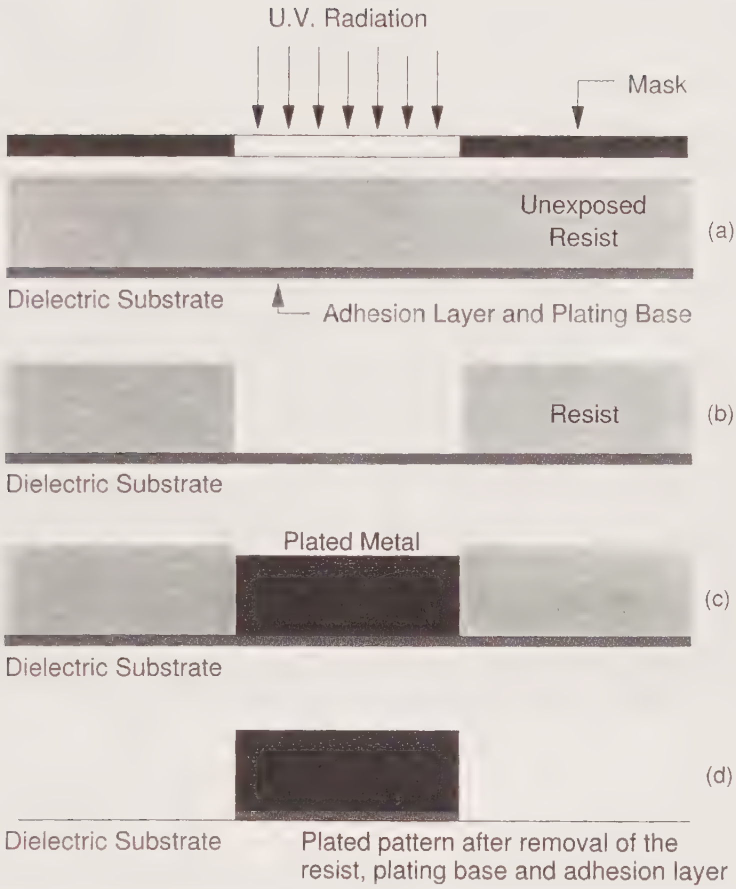

U.V. Radiation
Mask
Unexposed Resist
(a)
Dielectric Substrate
Adhesion Layer and Plating Base
Resist
(b)
Dielectric Substrate
Plated Metal
(c)
Dielectric Substrate
(d)
Dielectric Substrate
Plated pattern after removal of the resist, plating base and adhesion layer

Fig. 5.14. Schematic representation of the process steps in the plating-through-mask technology.

exposure and second plating thus obtaining two levels of plated metal using one resist layer and one seed layer. After plating the seed layer and adhesion layer are then removed by either chemical etching, sputter etching or reactive ion etching. This leaves the plated features isolated on top of the dielectric substrate. These features, when building a multi-layer structure, are then overcoated with a permanent dielectric. If necessary, the structure is planarized and the process may be repeated as many times as necessary to complete the device.

If a high resolution resist is used, the pattern can be very densely packed. When optical lithography was replaced by e-beam and by the x-ray lithography, pattern dimensions as small as 0.05μm were achieved.[55,56] An additional important feature of x-ray lithography is that, because of the extremely short wavelength and excellent collimation, it is capable of producing nearly vertical walls in very thick PMMA resist as shown in Fig. 5.15, which was published by Spiller et al.[55] in 1976. This permits fabrication of three-dimensional structures with vertical walls. It is this feature that makes plating through a mask so unique and attractive for fabrication of advanced electronic devices, and particularly in MEMS.

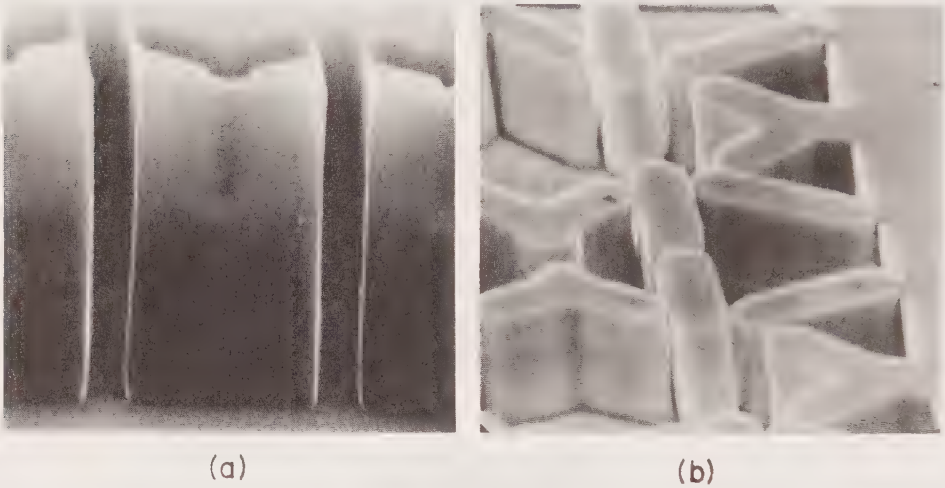


Fig. 5.15. (a) A 12 μm thick PMMA resist exposed in the DESY synchrotron in 1975 with a 1 μm bubble memory mask. Narrowest spaces are approximately 1,000 Å. (b) 6 μm of Au electroplated into the resist structures shown in (a), which exhibit resolution down to 1,000 Å. Mag. = 25,000X.[56]

5.5.4 Electroforming in Electronics is Known by Many Names

The application of electroforming in electronics is known under several different names: plating-through-lithographic-masks; plating-through-masks; pattern-plating; plate-up, and additive plating. Recently, the acronym *LIGA* has been appearing in the field of micro-electro-mechanical systems (MEMS). LIGA is an abbreviation of the German ***L***ithographie, ***G***alvanoformung and ***A***bformung (lithography, electroforming and replication). This name was coined by Kernforschungszeutrum Karlsruhe (KfK) in Germany.[70,71] In reality, "galvanoformung" translates into English as electroforming or plating, "lithographie" translates to lithographic mask formation; hence plating-through-mask technology is equivalent to

German *LIG*. The letter "A" in *LIGA* relates to "abformung" or injection molding. It is injection molding that is anticipated to result in mass production of millions of very small and inexpensive parts and components by injection molding of plastics, metals, or ceramics into hundreds of metal micromolds produced by synchrotron x-ray lithography in PMMA resist, followed by metal mold electroplating. Today, the mass mold replication of microparts produced by synchrotron radiation is in the early stages of development. So far, very few products are fabricated in this manner. Much work is needed to make it a practical and cost effective manufacturing process. Advanced development work in this area is carried out in the Karlsruhe Forshung Institute where the LIGA process originated. Most of the other LIGA oriented centers at this point in time are practicing only deep etch x-ray lithography, using synchrotron radiation, and electroplating.[72,73]

5.5.5 Patterning and Pattern Transfer

For a long time, the most popular technique for pattern transfer in electronics was wet chemical etching. In this process, metal is deposited as a blanket film, the portions of the film which are to be part of the final pattern are protected by a photoimaged polymer mask, and the remaining metal is removed with an appropriate etchant. The mask can be thought of as an umbrella which protects the pattern from being etched while the etching is done by spraying or vigorous agitation. The chemical etching process is actually an isotropic, accelerated corrosion process. Just as in corrosion, certain grains or intergranular regions will etch much faster than the others. The isotropic behavior and the attack of individual grains or intergranular regions result in a trapezoidal crossection of the conductor, with reduced crossectional area and very jagged edges which are hard to overcoat with a dielectric. The isotropic nature of chemical etching limits the closest spacing between conductors (patterns) to twice the film thickness. If the thickness of the film is 1μm, then the closest spacing that can be achieved is 2μm. That means the top of the conductors will be 2μm apart while the bottom will be just opening up (breaking the contact between the two lines). In reality in manufacturing this separation is approaching ~3X film thickness. Figure 5.16 shows aluminum and AlCu lines produced by chemical etching. This photo clearly indicates that aspect ratios of thickness to width cannot be greater than 0.5:1.0. When the top of the line becomes too narrow because of the need to etch narrow lines in a thick film the resist begins to break off due to the narrow area of contact.[3]

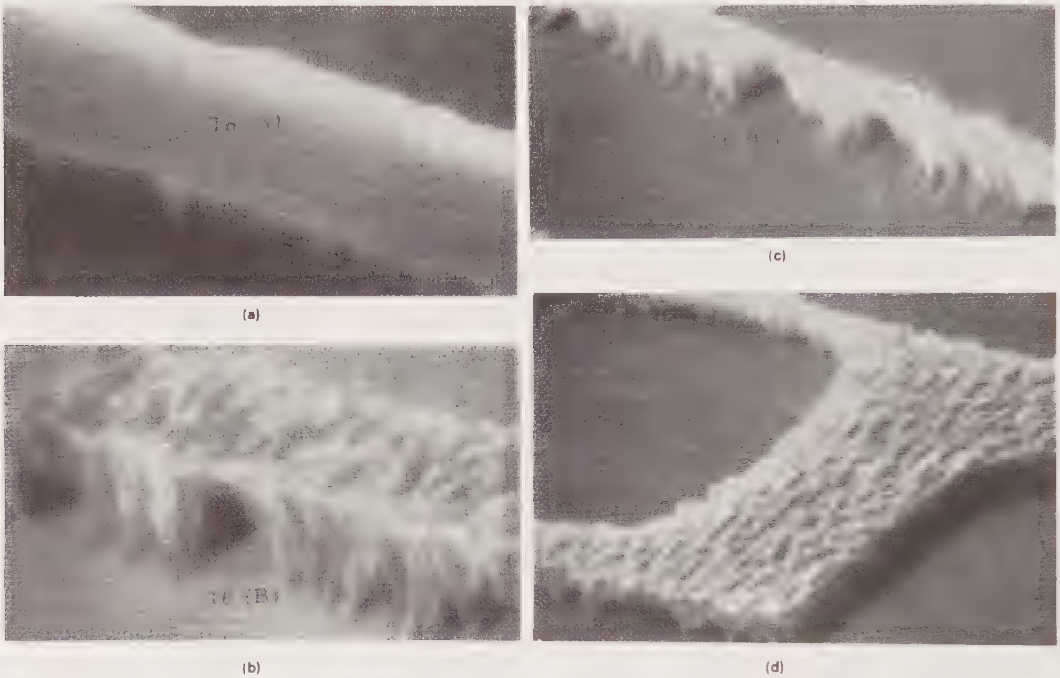


Fig. 5.16. (a) A 1 μm thick A1 film etched for 160 s using a nominally 5 μm wide mask opening using a typical phosphoric acid-nitric acid etchant. (b) Same as (a) but A1 film first heated for 2 hrs in vacuum. (c) Same as (b) but A1 film etched using a 2.5 μm wide feature mask. (d) A 10% Cu-A1 film heated as in (b) and etched for a shorter time.

The chemical etch process started approaching its limit when increasing device density made it necessary to place conductors closer and closer together while circuit requirements demanded that the largest possible cross-sectional conductor area be maintained. The limitations of chemical etching, which first became apparent in the circuitization of silicon devices and in the attempt to fabricate thin film inductive heads, prompted investigations of other techniques for patterning of thick conductors. Since electrical engineers and physicists are usually very familiar with vacuum techniques, it was only natural to see a considerable effort put by the electronic industry into dry-vacuum processing. The result was that sputter etching, RIE and lift-off became part of the electronics industry.

Sputter etched or RIE'd patterns can have aspect ratios of nearly 1.0. However, due to the nature of the process, the profiles are also trapezoidal, although the edges are closer to vertical. In sputter etching,RIE, or ion milling, a thick resist is usually applied, exposed and developed and that resist is used as mask. The energetic ions which are accelerated toward the part being etched, hit the edge of the resist and are deflected at an angle and continue toward the bottom of the pattern, eventually digging out a trench. This particle deflection continues throughout the total process. The resultant profile is shown in Fig. 5.17.

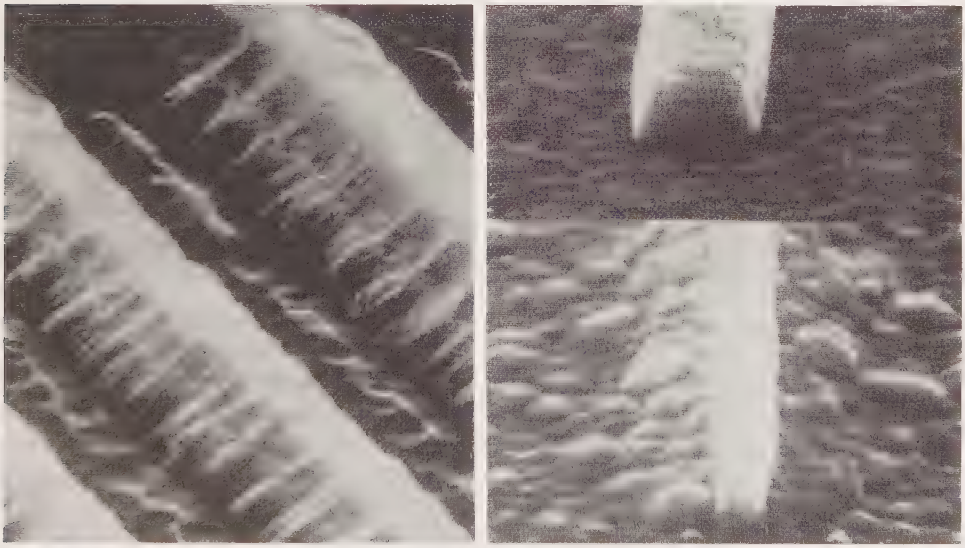


Fig. 5.17. Nominally 5 µm wide A1 lines sputter etched in an Ar plasma. Etching was stopped short of completion. The resist shows severe erosion. Indentations in the SiO_2 indicate a nonuniform etching rate for different grains and for the intergranular regions. (An A1 residue remained in the slower etching intergranular regions.)

In addition to the sloping wall, in sputter etching, RIE and even ion milling, the edges of the pattern are very often covered with a mixture of metal and resist polymer that tends to deposit from the plasma during processing. The disadvantage of this is that, (a) the redeposited material does not always adhere well to the metal, and (b) the redeposited material, in addition to metal and carbon, can also contain sulphur and chlorides which can be very detrimental to the structure fabrication process and reliability. These edge deposits, occasionally referred to as "fences," are considered to be a serious problem. The chemicals necessary to remove them are usually very harsh. From the point of view of MEMS, RIE, sputter etching and ion milling do not permit one to readily achieve the tall structures with narrow vertical spaces that MEMS usually requires.

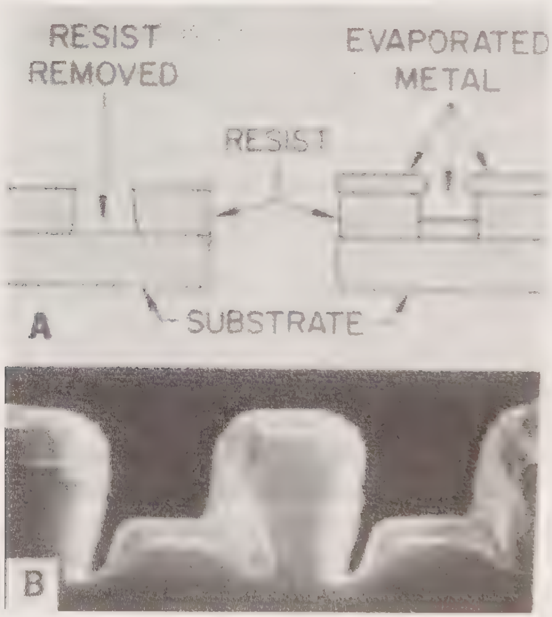


Fig. 5.18. (a) Schematic representation of the lift-off process. (b) Cross-section of metal lines and metal-coated resist prior to lift-off.

The lift-off process (Fig. 5.18) has been developed primarily for the fabrication of AlCu conductors on silicon chips to provide the wiring interconnect pattern. The highest practical aspect ratio achievable in this process is ~1:1. The limiting factor here is the production of the T-shaped pattern with a narrow foot and wide overhang; this also limits the line density. The metal can be deposited only by line-of-sight processes such as evaporation from a distant source so as to avoid the angle of incidence issue which results in coverage of resist under the T with thin layer of metal and makes lift-off hard to accomplish. It is for the same reason that plasma deposition processes cannot be used in association with lift-off. The metal thickness has to terminate before the overhang is reached. This is because the solvent must get under the overhang in the final step of the process to dissolve the resist and thereby lift off the excess metal. The resulting conductor is slightly rounded as shown in Fig. 5.18.

The limitations of chemical etching and the shortcomings of the various dry process patterning techniques in meeting the needs of thin film magnetic head fabrication led to the effort at IBM in the 1970s to develop through-mask plating. A study published in 1976[3] comparing the patterning of metals by wet chemical etching, sputter etching, lift-off and through masks electroplating shows the capabilities of the latter technique. A pictorial comparison of the various patterning processes is shown in Fig. 5.19. It is easy to conclude from this figure, that for the same pattern width and using the same resist mask, the plating-through-mask process is the one technique that permits the highest packing density, the highest cross-sectional area and the tallest structures. These aspects will be discussed in more detail later in this chapter. This technology has been clearly demonstrated in electronics to be the only one that is capable of producing truly three-dimensional, micron-size structures excellent resolution and vertical walls. It assures extendibility of a wide variety of devices in electronics and opens up many possibilities in MEMS.

5.5.6 Beginnings of Plating Through Mask Technology in Electronics

In electronics, the plating through lithographic polymeric mask technology, or electroforming, began in the late 1960s. Several inventions and breakthroughs discussed in the literature[3,10,47–62] led to the current extensive use and popularity of the electroforming process in electronics. Electroforming through masks was first tried in connection with fabrication of magnetic storage devices.[50–52] In order to build thin film inductive read-write heads, it was necessary to produce micron-size copper conductor coils which were nearly square in cross-section, with as many conductor turns crowded around the magnetic yoke as possible.[51] As seen in Fig. 5.20, the yoke was only 100μm high, with a separation between the yoke legs of only 15 to 20μm. The structure is shown compared with

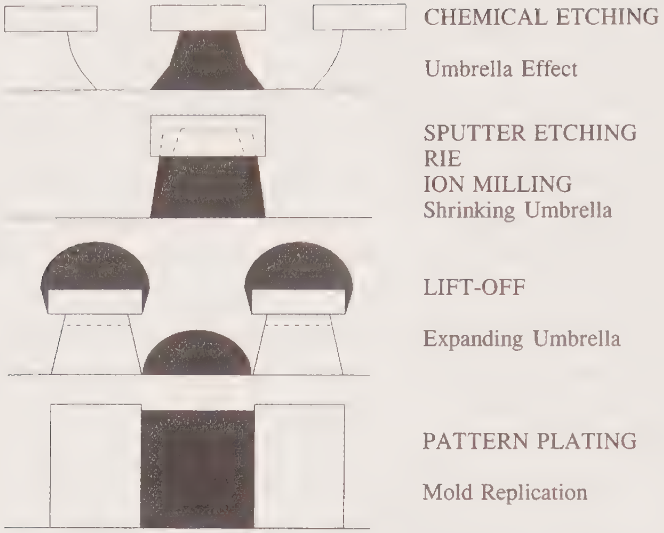
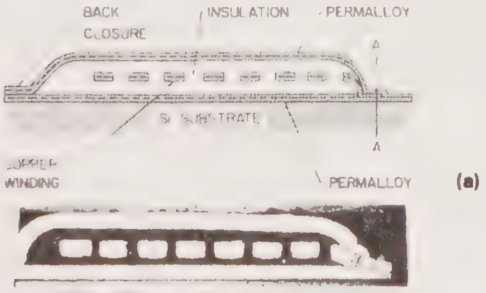


Fig. 5.19. A schematic summary comparing four pattern transfer techniques — chemical etching, sputter etching, reactive ion etching, lift-off, and plating through mask technology— using UV lithography and the same line-width mask and the same line spaces. Because of the nature of the plating process, even the smallest features that can be wet by the plating solution should plate. The height-to-width aspect ratio of the pattern depends only on how the lithographic mask can be made.

a crossection of a human hair.[64] The copper coils were laid out around the magnetic horse-shoe in a pancake fashion after the first leg of the horse-shoe magnet was fabricated. The permalloy was also plated through a resist mask. The "resist frame" was used to define the outline of the yokes and of the pole tips.[57]

The fabrication process exploited, for this purpose, the combination of evaporation, sputtering, lithography, electroplating, and chemical etching techniques. The individual layers of the multilayer structure were formed by sequentially alternating metal and the dielectric layers. For fabrication details, the reader is referred to numerous papers, patents[51,52,58-65], and a book chapter.[9] Considering the 1969 through 1974 time frame in the historical development of electronic and magnetic device fabrication, this was perhaps the most advanced structure fabricated. Every step in the process had to be invented or adapted, and thoroughly tested. The process integration required all process steps, all materials, and all chemicals to be compatible with each other. The number of process steps was greater than in fabrication of bipolar devices being actively developed at approximately the same time. By comparison with semiconductive devices, the individual layers were much thicker. The magnetic layer had to have very smooth surfaces, and sharply defined edges for the head to be functional. Many of these requirements were discussed in Ref. 9.

The "resist frame plating" scheme shown in Fig. 5.21 represents another variation of the through-mask-plating technique. Such resist frame plating solved not only the Permalloy patterning challenge, but provided a well defined structure with smooth vertical walls, which were essential to magnetic functioning of the heads. It also served two additional functions: it minimized the plating current density variation over the wafer by minimizing the masked regions, and it provided a very uniform current distribution, and hence very uniform Permalloy thickness and composition; it was also making it possible to achieve good magnetic anisotropy in Permalloy with application of only 500 Oe of applied magnetic field rather than 10,000 Oe, which would have been necessary to overcome the demagnetizing field if yokes and pole tips were electroplated as discrete patterns.



Early version of a thin film head-Yorktown

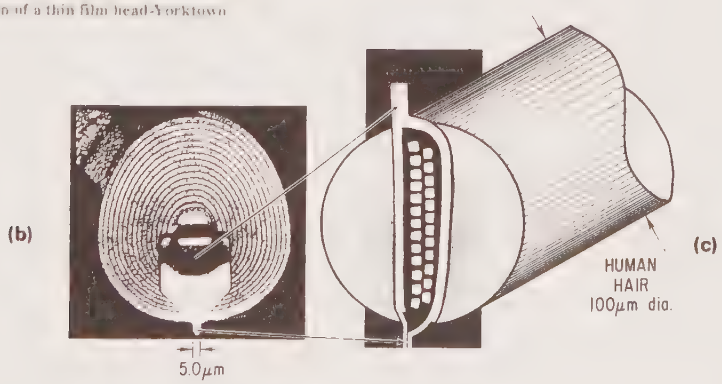


Fig. 5.20. (a) Cross-sectional area of an early experimental version of a vertical head. (b) and (c): cross-sectional area of a manufactured inductive head of recent vintage and a comparison of the cross-sectional areas of the latter with a human hair.

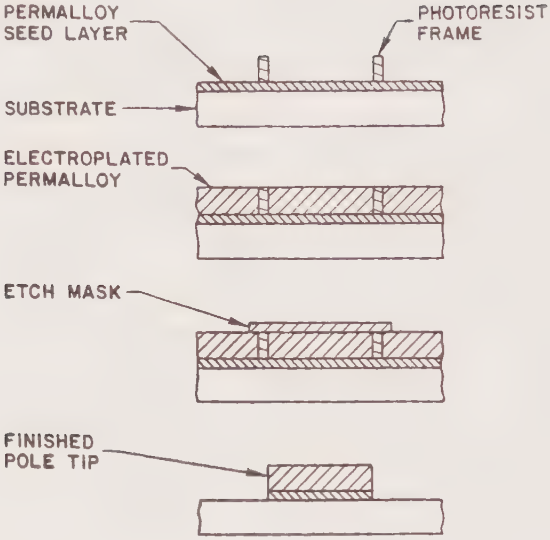


Fig. 5.21. Schematic representation of frame plating. The photoresist frame patterns perform four separate functions: (a) they permit uniform current distribution, hence uniform composition and thickness of the plated film, (b) they achieve well defined patterning, (c) they afford smooth and well defined Permalloy walls, and (d) they allow the use of a relatively low magnetic field to achieve good magnetic anisotropy in the pole tips.

After etching the background Permalloy and using Al_2O_3 as a gap material, the copper coils were imbedded into hard baked Novolac positive working resist as a dielectric. Novolac resist flows when baked to above 200°C, thereby minimizing its surface area due to the resulting gradually sloping and rounded walls. Upon cooling, the hard-baked Novolac resist sets and becomes an insoluble dielectric. The pancake-shaped copper conductor coils were formed by plating the 3 to 4μm thick, ~3μm wide copper coils onto an 800Å thick copper seed layer through a Novolac positive working resist. After formation of the conductors, the resist was dissolved and then the seed layer was removed by ion milling or chemical etching. Since the seed layer was only 800Å thick, it was possible to remove it by ion milling or even chemical etching without employing a mask, while removing only a negligible amount of copper from the top of the electroplated conductor in the process. The Ti adhesion layer was removed by a short chemical etch in 0.1% HF solution in water.

In the currently manufactured heads, once the first coil is formed, as described above, a new layer of Novolac resist is applied and hard baked so that it flows into the spaces between the conductors and serves as an additional dielectric. After further levels of coils are formed by a plating-through-mask process, the second part of the 2-5μm thick Permalloy yoke is again plated through "frame masks" onto a thin sputtered permalloy seed layer. The fabrication of the second part of the yoke is completed by selective permalloy etching as has already been described for the first leg of the yoke. For more details about the electroplating and fabrication process the reader is referred to an historic overview of the thin film head process development[51], and to a recent book chapter by Andricacos and Romankiw.[9] Many attempts were made by others to build the inductive thin film heads by chemical or sputter etching of the conductor coils and Permalloy yoke. The conductor spacing and thickness needed proved to be beyond the capability of these techniques.

The magnetically functional inductive thin film heads described here, could not be built efficiently and cost effectively by patterning techniques other than plating-through-mask technology. The switch from ferrite to thin film heads represented a complete shift of the paradigm. This switch came in several years earlier than it otherwise would have if plating though resist masks had not been developed. The use of thin films and the plating through-resist-mask process permitted for the last 18 years an increase in storage areal density by one order of magnitude every 8 years.

The thin film heads were the first devices in which the full power of the plating-through-mask technology was demonstrated. It was particularly effective along with the other major breakthroughs and advances which were made at the same time, such as: development of current density insensitive Permalloy baths[58,59]; understanding how to plate copper coils of uniform thickness through a resist mask despite wide separation between individual heads; use of the hard baked Novolac resist as a dielectric[60]; seed layer removal from between tall structures with narrow spaces, and development of completely new plating tools and methods of agitation.[61] Due to the lack of availability of viscous, thick-layer-forming Novolac resists, a method of spincoating 5 to 15μm thick Novolac resist layers was developed using formulations which were aimed at spinning 1 to 2μm thick layers. This was achieved by using a long resist dwell time prior to spinning, and by developing a multiple-layer spinning technique.[62]

As the understanding of the plating through resist masks was improving, precision tools, processes and process controls were developed, and through-mask plating technology became a critical technology in magnetic thin film head manufacturing. Today, several hundred million heads per year are manufactured by electronic companies around the world, using the plating through mask technology, representing a $2-3 billion market. There is currently no deposition method that can compete costwise with electroforming for thin film head manufacturing.

5.5.7 Bubble Memory, E-Beam, and X-Ray Lithography

In parallel with the introduction of the electroforming through resist mask process in thin film head manufacture, it was also being introduced in the fabrication of bubble memory.[52-55] Because bubble memory required submicron dimensions, the electroplating process developed for optically exposed resist patterns was extended to e-beam[54] and x-ray lithographically formed patterns.[55,56]

The first results of the plating-through-mask technology in connection with e-beam lithography were reported by Romankiw, Hatzakis and Castellani in 1974 at the Spring ECS Meeting.[54] At that time e-beam lithography was in its early stages of demonstration and development. Figures 5.22(a) and 5.22(b) show the smallest gold conductors and gold studs achieved on plated Permalloy seed layers. They were 0.6μm tall, 0.3μm wide at the base, and 0.6μm wide at the top. Plating was done through a mask created in PMMA resist using electron beam writing. The e-beam technology of exposure in PMMA along with electroplating of gold through the PMMA resist mask, was also successfully used in our lab to create zone plates with gold features down to 700Å in size, with 300Å separations between the features.

High resolution x-ray masks require thick gold patterns on x-ray transparent, thin boron-doped silicon membranes. Because of the submicron dimension required for the bubble memory device, such patterns had to be written by e-beam in PMMA resist. Initial investigations were started using evaporated films of Au which were ion milled using a PMMA mask. Due to redeposition, the edges of the Au pattern were sloped, and trenching due to deflection of energetic ions was weakening or fracturing the thin membranes.

The lift-off method that was initially employed in this work yielded sloped edges and, therefore, equally fuzzy edge definition in the final

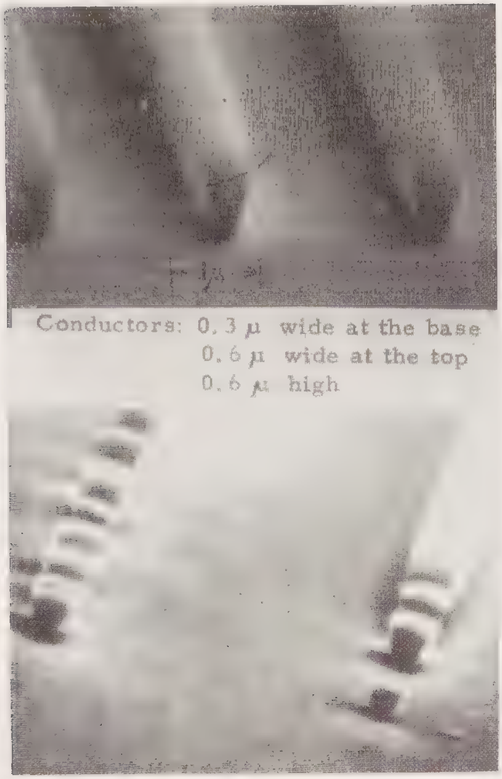


Fig. 5.22. Small Au conductors (bottom) and studs (top) plated on Permalloy seed layers using electron beam exposed PMMA.[54]

x-ray mask. This prompted investigation of electrodeposition as an alternate method. The boron doped silicon membranes were metallized with 50Å Ti adhesion layer and 300Å of gold conducting seed layer prior to application of thin layer of PMMA. After exposing a pattern in the PMMA, a thin gold pattern was electroplated into the e-beam generated pattern. This mask was used to do x-ray exposure into much thicker PMMA, and thicker Au masks were then electroplated. The 50Å of Ti and 300Å of Au seed layers were thin by comparison with the microns-thick Au patterns, and therefore they did not need to be removed for the mask to be used. The electroplated patterns, by comparison with ion beam milled and with lift-off Au patterns, gave very well defined features with vertical walls, thus resulting in well-defined, sharp pattern edges. Each subsequent mask inversion permitted the production of thicker gold features with equally excellent edge definition, and almost the same fidelity as was observed in the first e-beam generated mask.

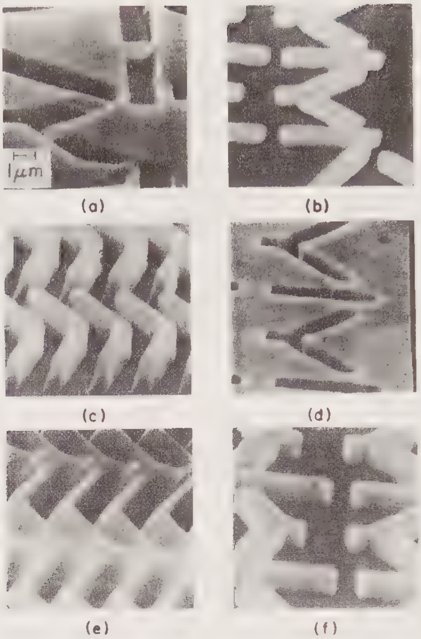


Fig. 5.23. The fabrication steps for a bubble memory using Au electroplating for all metallization steps. Left are X-ray resist patterns, and right are metal patterns after resist removal. The e-beam fabricated master mask is on top, the negative X-ray copy is in the middle, and the device pattern obtained from the negative mask is on the bottom. Line widths are 1 μm.[55]

The gold patterns in Fig. 5.23 that were plated into e-beam exposed resist were inverted twice and exhibit excellent fidelity of reproduction of the original mask.[55] Based on this work, electroplating was adopted as the standard method of x-ray mask fabrication, and is still the favored technique for x-ray mask fabrication today — 23 years later. High aspect ratio structures with up to 12:1 were produced, not only using x-ray lithography, but also using optical lithography and through-mask plating (Fig. 5.23). The optical lithography results shown in Fig. 5.24 are excellent and are presented here as a possible alternative to x-ray processing when access to the latter technique is not possible.

5.5.8 Computer on a Chip

In 1973-74 a study was conducted in which an experimental pattern of 2μm-wide conductors was used with progressively narrower spaces ranging from 1μm down to 0.3μm, which was built using Novolac posi-

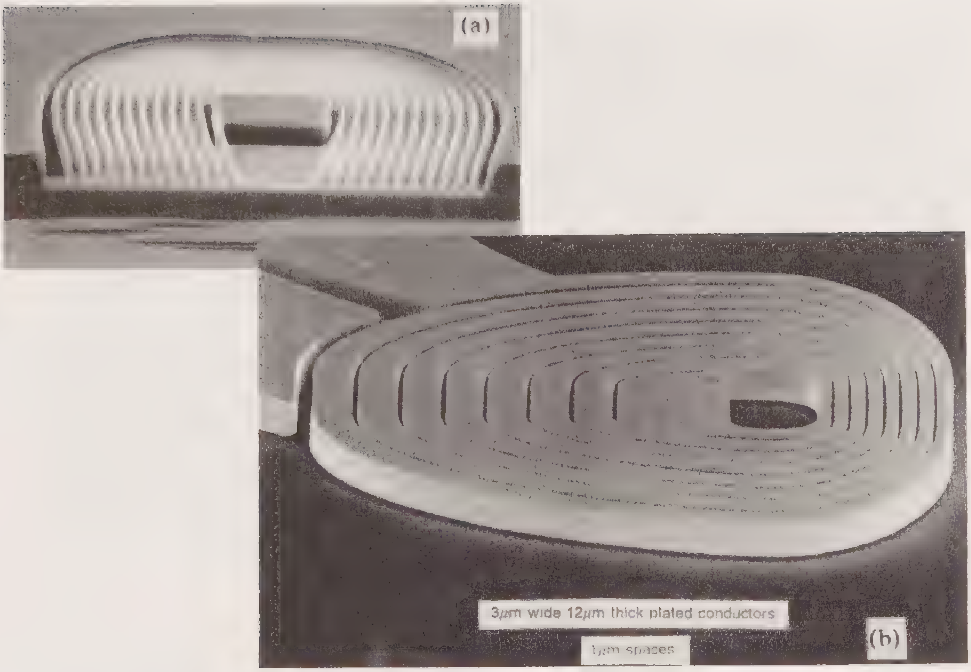


Fig. 5.24. (a) Patterned 12 μm thick Novolac resist. (b) A12 μm tall, 3 μm wide coil of Au electroplated through the patterned resist.

tive working resist, A five-level structure consisting of three conductor levels and two via metallization levels was constructed. The structure consisted of 0.8μm thick, 2.0μm wide gold or copper conductors interconnected by 0.8μm diameter studs. As dielectric, hard baked Novolac resist and polyimide were used.

The objective of this project was to give an indication if it was possible to do a considerable amount of metal interconnect packaging on silicon, and perhaps in the future to build a "computer on a chip". The results of this study were reported at the ECS Fall Meeting in 1974.[66] The contact resistances were low, and wires between all levels showed good connectivity (Fig. 5.25).

5.5.9 Multi-Chip Thin Film Package

Figures 5.26(a) and 5.26(b) show portions of a thin film, multi-chip package consisting of X and Y planes, two ground planes and top surface pad metallurgy — a total structure consisting of 11 metal layers of

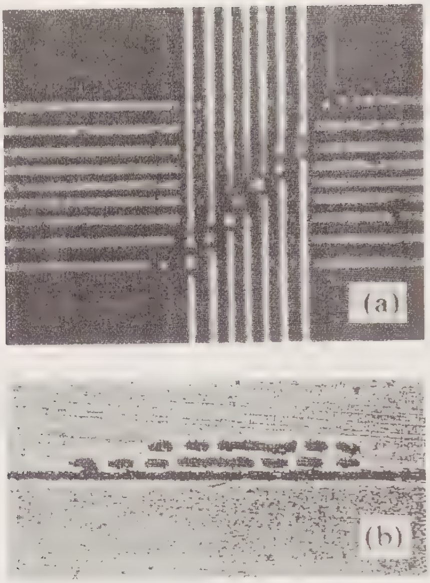


Fig. 5.25. A five-level structure consisting of three conductor levels and two via metallization levels fabricated in the early 70's. The structure consisted of 0.8 μm thick, 2.0 μm wide Au or Cu conductors interconnected by 0.8 μm diameter studs.

10μm wide and 5μm thick conductors, which were successfully built by plating-through-mask technology using polyimide as a dielectric in the 1989-91 period.[67] In the SEM photos in Figs. 5.26(a) and 5.26(b) the wiring is shown with polyimide removed from a portion of the structure by plasma treatment to reveal the metal wiring structure and to check for integrity of the interconnects. Elements of this technology are now used in manufacture of thin film multi-chip carriers. Subsequently, through-mask-electroplating was investigated also for interconnection with PbSn plating to electrochemically fabricate mechanically robust C4 interconnects between the silicon chip and the thin film package.[68]

5.5.10 High Precision and Tight Tolerances - New Equipment Needed

Electronic parts require very tight thickness tolerances, exceptionally uniform current distribution and mass transport, rapid real-time on-line process control and a good understanding of the relation between solution chemistry, nucleation, film growth and other parameters that define the

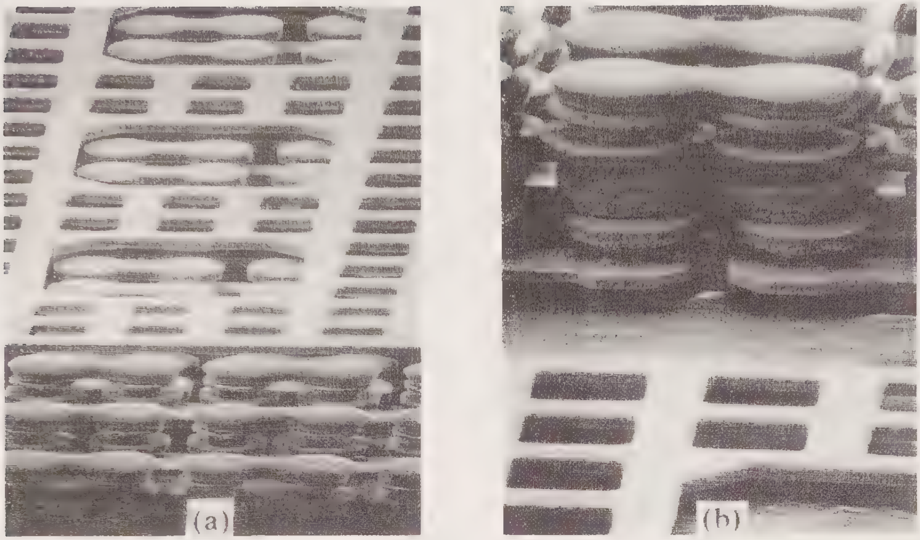


Fig. 5.26. 10 μm wide, 5 μm thick copper conductors formed in an X-Y plane surrounded on each side by a ground plane and top surface metallurgy. The structure was built to demonstrate the ability to build thick-film multichip carriers. The polyimide dielectric was removed to permit better viewing of the structure.[67]

material, magnetic, electrical and other properties of electrodeposits. Most of these requirements are affected by uniformity of agitation, pattern distribution and current distribution. A unique paddle plating cell design was developed which provides extremely reproducible agitation, and permits plating[61], very reproducibly, films with excellent current and thickness distribution over an entire wafer.[45,46] It is well known that agitation in plating tanks is hard to reproduce, and that scaling up of the plating environment to achieve the same diffusion layer thickness over entire cathode is very difficult. In the paddle cell, because of the unique method of agitation used, agitation is easy to reproduce and the paddle plating tanks are easy to scale up from a small laboratory size tank to a full-scale manufacturing tank. The agitation is accomplished by a windshield-wiper-like device (arm) moving only a millimeter away from the cathode. The paddle which sweeps the entire width and the entire length of the cathode at a frequency of 1 Hz assures very uniform diffusion layer thickness over the entire cathode and permits easy scale up.

The agitation at 1Hz frequency is equivalent to a diffusion layer thickness achieved in a rotating disk electrode operating at 400 rpm. Such a level of agitation would be extremely difficult to achieve by solution recirculation, impingement, back and forth movement of the cathode, and at the same time to be able to maintain the same degree of diffusion layer thickness uniformity which is achieved by the paddle movement at close proximity to the cathode. The agitation at the close proximity to the cathode assures uniform supply of ions (critical in NiFe alloy plating) and assures H_2 gas bubble release from the recesses in the resist pattern. Additionally because of the mechanical paddle movement very little plating solution has to be recirculated by pumping. Figure 5.27 shows a schematic drawing of a paddle cell. By investigating relationships a model was arrived at which permits to approach the electrochemical process with a considerable understanding.[42,45,46]

5.6 LIGA AND HI-MEMS

According to the historical account by Wolfgang Ehrfeld[69], electrodeposition specifically oriented to MEMS was initiated in early 1980. It was first used to build an isotope separator which required tall and small structures[70,71]. In the application and development of LIGA to MEMS one can distinguish three different levels of development of this technology: a) single layer structures, single exposure; b) single layer exposure with part or the whole structure being released; c) multilayer x-ray and optically exposed structures with several optical and/or x-ray mask alignments.

5.6.1 Single layer structure

X-ray radiation from synchrotron was used to expose and develop thick PMMA applied on top of conducting surface. After pattern development the thus-formed mask was filled with metal by electroplating of nickel. At first it was envisioned that thus-electroformed microstructures could be used directly or after one more inversion as molds to form a micropart by injecting molding into the nickel metal mold: plastics, metals, ceramics, composites, etc. In these situations only one layer of electroplated metal is used. This technology is being developed further by KfK. Such molds are used to produce optical microparts, lenses, deflectors, and spectrophotometers.[72-75]

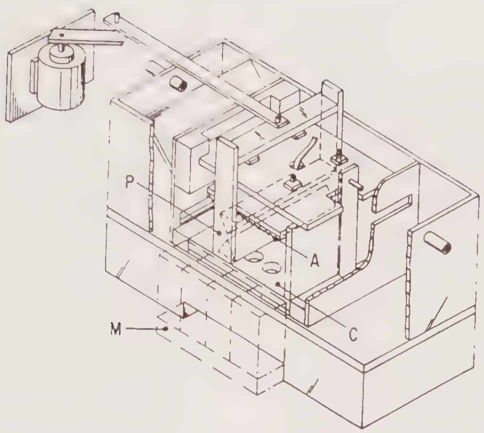


Fig. 5.27. Schematic representation of a paddle cell; A = anode; C = cathode; P = paddle; M = magnets. A detailed description of the cell is found in the literature.[43]

5.6.2 Sacrificial Layers and Release of Electroplated Components

The second stage in development came when it was discovered that it is possible to release either the whole or part of the electroplated structure. The combination of sacrificial release layers and electrodeposition through thick resists is a powerful technique for the formation of the important class of *movable* micro-components such as cantilevers, gears, and motor rotors, and other structures such as bridges and membranes. Metals such as Cu, Zn and Cr can be used as sacrificial layers in conjunction with electroplated structural metals such as Ni, permalloy and Au. Polymer release layers may also be used in certain circumstances. Even though a potential sacrificial layer material may appear to exhibit similar reactivity to the metal, which is required to be electroplated, it may be possible to identify an etchant which is selective to the sacrificial layer only.

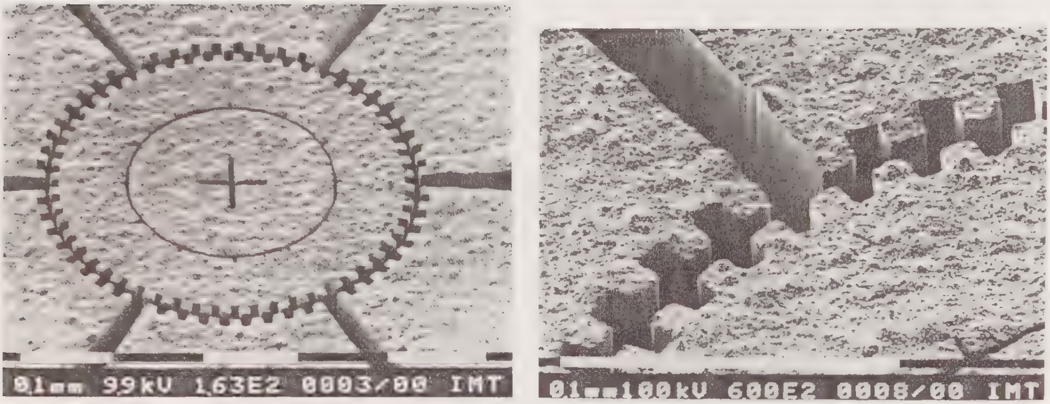
There are a large number of such devices described in the literature. Many of them are only in the feasibility of development stages. Several papers represent the early classical LIGA structures using one release layer and one electroformed metal layer to produce movable parts (Wallrabe et al.[76]). These authors demonstrate the effectiveness of use of the sacrificial layer technique which allows one to produce movable structures. They use the technique to build and operate an electrostatic stepping motor. The motor was built by electroplating 100μm of nickel into a mold in which the rotor has a release layer built in underneath. Figures 5.28(a) and 5.28(b) show the general view of the motor and the details of the structure.

Guckel et al.[77] have used the sacrificial layer technique and fabricated by plating through an x-ray exposed mask a planar magnetic motor and gears. In their case some of the rotating parts were built on a separate substrate and were then fitted over the posts, which were produced on a substrate without a sacrificial layer. Their assembled structure is shown in Fig. 5.29. The nickel was plated to a 100μm thickness. The motor is claimed to have been driven at up to 8000 RPM and this speed was maintained for several days. The motor shown in the figure has overcome the static friction at ~40 g. The nickel was reported to have very excellent wear resistant properties.

Rogner et al.[78] see applications of movable fabrication techniques as useful in the field of integrated optics — especially for the precisely positioned stop faces and precisely located spring elements. They fabricate a variety of fiber chip coupling structures, metallic gear wheels and capacitive acceleration sensors. Their paper explains in detail how to build a structure and how to release it.

Harmening et al.[79] describe a method for the production of stepped structures. They also provide a detailed explanation of how to produce such structures using a combination of LIGA, electroforming of molds and then forming a second level structure on top using a second lithography level. This paper is interesting from the point of view that it uses a second alignment and exposure step. The discrepancies between the first and the second step in lithography are compensated by dimensional allowances in fabrication of the second x-ray mask.

The next two devices reported by the Delco Electronic Devices[80,81] are now in a product program. Consumer demand for increased automotive electronics and safety features has created a market for low cost automotive angular rate sensors. Delco has recently combined electroformed metal parts using LIGA with CMOS technology to produce an angular rate sensor and a CMOS integrated, electroformed accelerometer. Potential applications are yaw control, rollover detection and turn-by-turn navigation. In the angular rate sensor the vibration of a partially released ring 1mm in diameter is capacitively monitored and amplified by a



(a) (b)

Fig. 5.28. (a) SEM photograph of an electrostatic LIGA-motor made of electroplated Ni with 9 teeth on each stator. Stator radius: 267 μm, axle radius: 125 μm, bearing clearance: 4.8 μm, height: about 100 μm. (b) Close-up view of the motor at two adjacent stator poles. Tooth width on outer rotor radius is 14 μm, tooth distance is 10 μm.

CMOS device. The advances in micromachining by electroforming are providing the breakthrough necessary to make angular rate applications practical for the whole line of cars rather than only for luxury cars. Several hundred of these devices are produced on a wafer. Similarly, several hundred accelerometers are produced on a wafer. The electroformed angular rate sensor — a gyroscope — is used to indicate when the automobile is turning or swerving. They are particularly effective to prevent swerve on ice. They are much more sensitive than mechanically built devices, and greatly improve safety. Another application is to assist GPS system in automotive navigation.

In the accelerometer device the capacitance changes as acceleration takes place. These devices might eventually replace the silicon sensors that control the release of air bags. Eventually the two devices may be made on a single chip. Figure 5.30 shows the angular rate sensor and the electroformed accelerometer on a CMOS chip.

Another example may be micro electro mechanical tunable filter for infrared spectroscopy, reported by Ohnstein et al.[82] This work was done jointly between Honeywell Inc. and University of Wisconsin.[83] The released part is a tunable filter which consists of a multiplicity of x-ray exposed and nickel plated shutters/springs which are closed or opened by pushing a magnetic plunger into or out of a magnet. The shutters are

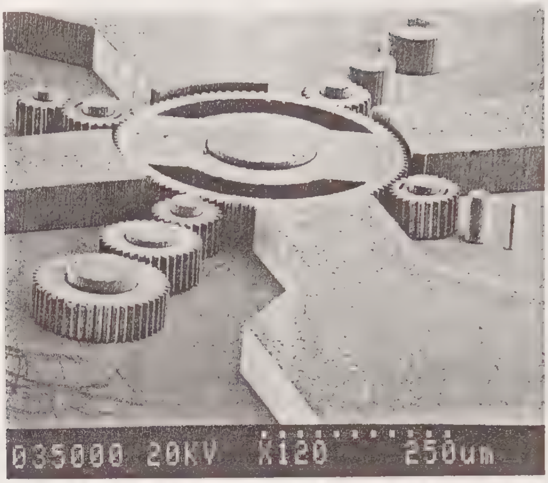


Fig. 5.29. An assembled reluctance motor fabricated using Ni electrodeposition by Guckel et al.[77] The Ni is 100 μm thick.

3μm wide, 50μm high and 6mm long. Most of the parts are made of nickel or Permalloy. The magnetic force is applied by using several copper wire hand-wound Permalloy C sections. This is one way in which H. Guckel of the University of Wisconsin fabricates a variety of magnetic linear and rotary actuators which may require a large number of coil turns.

Figures 5.31(a) and 5.31(b) show the overall layout of the tunable filter and the details of the actual shutter. Figure 5.32 shows a rotor with several hand turned coils arranged on a circle around the rotor. To produce such a motor Klein et al.[84] use one x-ray exposure — the plated Permalloy in several separate layers, polished each layer, and then stacked the layers together and aligned them using a peg assembly. The magnetic drive coils are produced by electroplating the horseshoe-shaped magnetic core into a very deep x-ray produced pattern in PMMA on a release layer. Following chemical release, the cores are then *hand-wound* with insulated fine wire conductors, and are used along with a rotor and bearing system to create quite high-performance minimotors. A possible disadvantage of this approach is the rather slow, delicate process of winding the coils and the delicate assembly process, both of which are not very amenable to mass manufacturing.

Allen et al.[85] used Cr as a release layer for Cu structures. In general it is advantageous to use sacrificial layers of one or more microns in

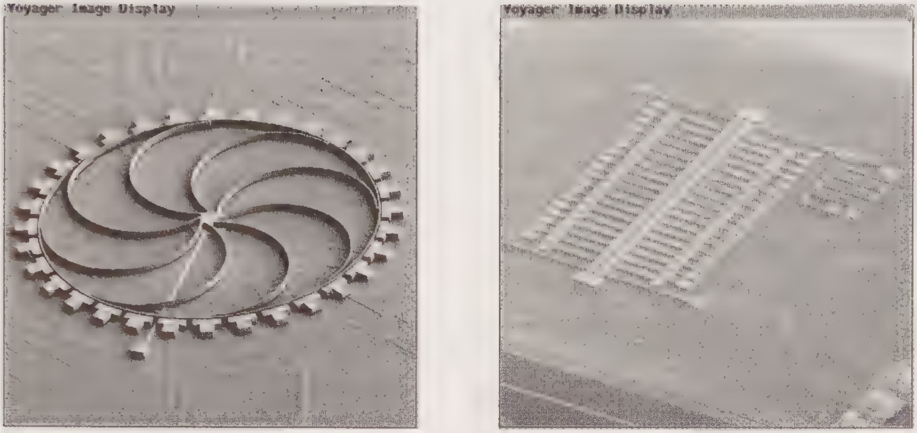


Fig. 5.30. (a) Photomicrograph of a 1 mm diameter electroformed angular rate sensor formed on top of a CMOS IC. The ring vibrates under vacuum. (b) A CMOS integrated electroformed accelerometer.[80,81]

thickness, if possible, to speed up the release process, and to minimize the risk of chemical attacking the structure being released, such as forming pits. The sacrificial layer may be electroplated in certain instances. Further discussion on sacrificial release layers may be obtained in the paper of Maciossek et al.[86]

Ahn and Allen[85] have described the fabrication of an innovative, *fully integrated* magnetic microactuator, which is shown in Fig. 5.33. In this structure, plated copper conductors located in one plane (i.e., horizontally) were interwoven with multilevel Permalloy "meander" magnetic cores. The magnetic microactuator part was a Permalloy cantilever beam, the beam part having been released using either 7μm thick Al or Cu sacrificial layers. The complete fabrication scheme is shown in Fig. 5.34. The cantilever beam was 2.5μm thick, 25μm wide and 780μm long; the meander-type coils contained 17 turns. The cantilever tip exhibited a deflection of 6μm in the vertical direction when a dc voltage less than 1 V (giving a drive current of 800mA) was applied to the coils.

Figure 5.35 shows a portion of an electrostatic, comb-drive-type linear actuator fabricated of Ni by Borner et al.[86] The Ni was electrodeposited through PMMA resist to a thickness of 80μm, and the resist was removed; the smallest feature is about 1μm. The aspect ratio in the original resist for this part was 100 for the small features. That such high-aspect-ratio features could be successfully developed out in the PMMA resist was

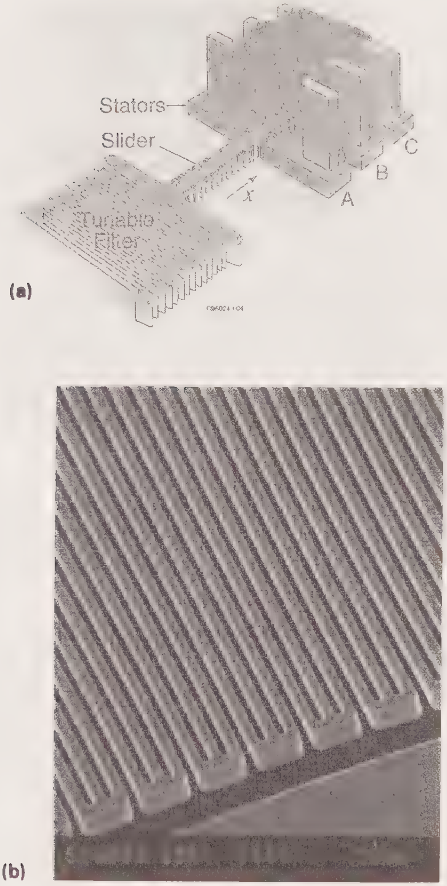


Fig. 5.31. (a) Tunable IR filter with integral stepping motor. The slider and tunable filter are electroplated from the same LIGA mask assembled onto the stator and mounting posts fixed to the substrate. The electromagnets are assembled into the stators after the coils are wound. (b) The fine, metallized gratings, which involved thin PMMA structures several millimeters long, were possible only by having the PMMA completely free of stress. The PMMA structures for the mold of this pattern were unsupported at both ends and are 3 μm in width, 6 millimeters in length, and 50 μm in height.

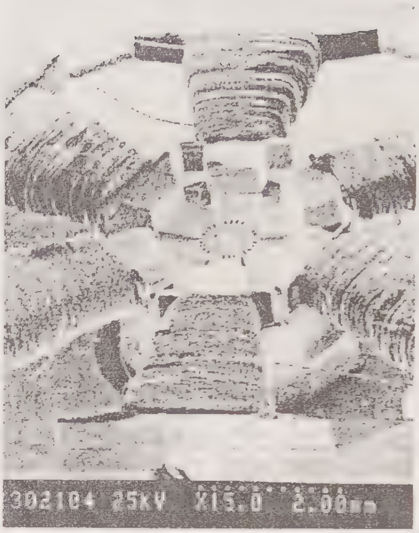


Fig. 5.32. SEM of a microinduction motor. The coils are made of three, 200 μm thick layers stacked and pegged together. The stator poles and rotor may also be stacked to achieve an arbitrary height.[84]

probably due to their being connected to larger open areas, which facilitated solution access by capillary forces.

5.7 EXAMPLES OF METAL ELECTRODEPOSITS

Figure 5.36 is an SEM picture of studs electroplated at IBM through PMMA resist with Permalloy (80at% Ni) to a height of ca. 150μm. The

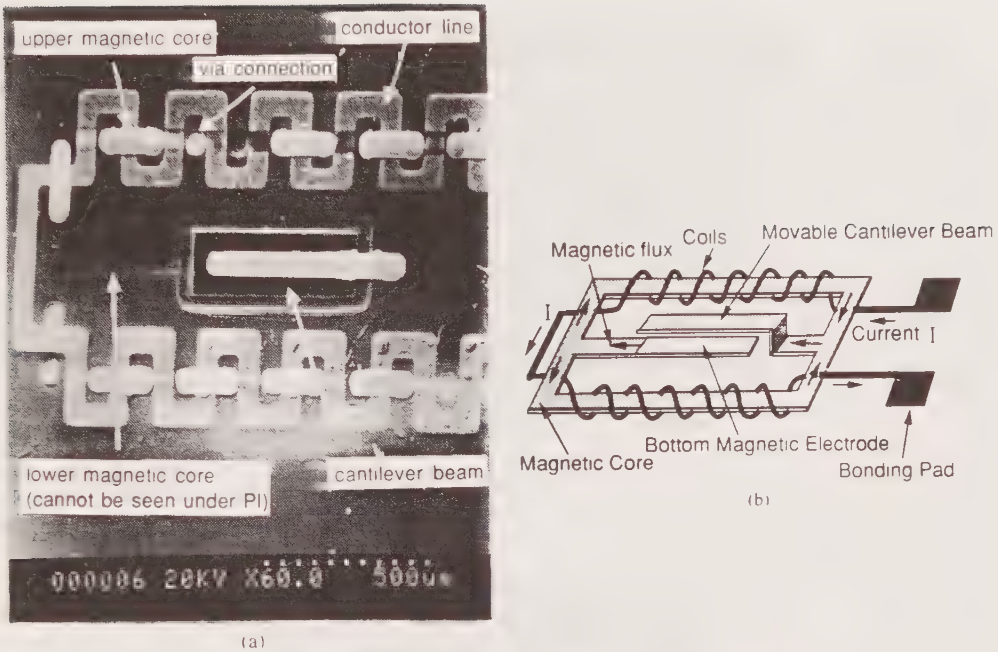

upper magnetic core
via connection
conductor line
lower magnetic core
(cannot be seen under PI)
cantilever beam
000086 20KV X60.0 500um
(a)
Magnetic flux
Coils
Movable Cantilever Beam
I
Current I
Bottom Magnetic Electrode
Magnetic Core
Bonding Pad
(b)

Fig. 5.33. SEM and schematic diagram of the integrated magnetic microactuator with nickel-iron cantilever beam. (a) Magnetic flux is guided both to the cantilever beam and to the bottom magnetic contact. Some upper magnetic core sections have been removed to show the magnetic vias. (b) Analogous schematic diagram: the actual meander type coil in the structure is represented as being equivalent to a "wrapped" solenoid type coil.[85]

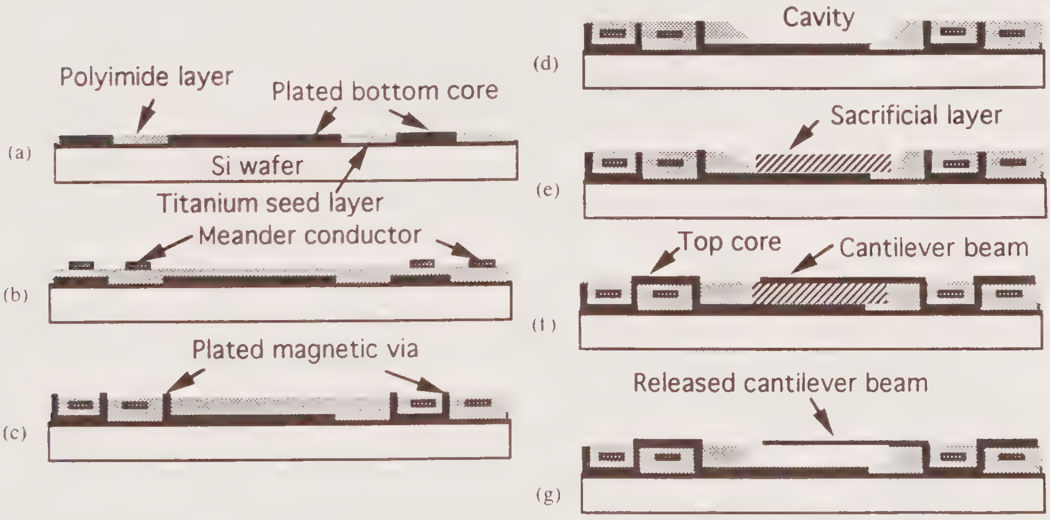

(a)
Polyimide layer
Plated bottom core
Si wafer
Titanium seed layer
(b)
Meander conductor
(c)
Plated magnetic via
(d)
Cavity
(e)
Sacrificial layer
(f)
Top core
Cantilever beam
(g)
Released cantilever beam

Fig. 5.34. Processing sequence used to build a fully integrated magnetic microactuator. (a) multi-spincoating of polyimide, dry etching, and bottom core plating, (b) patterning of conductor, (c) magnetic via plating, (d) dry etching of cavity, (e) patterning of sacrificial layer, (f) top core and cantilever beam plating, (g) release of cantilever beam.[85]

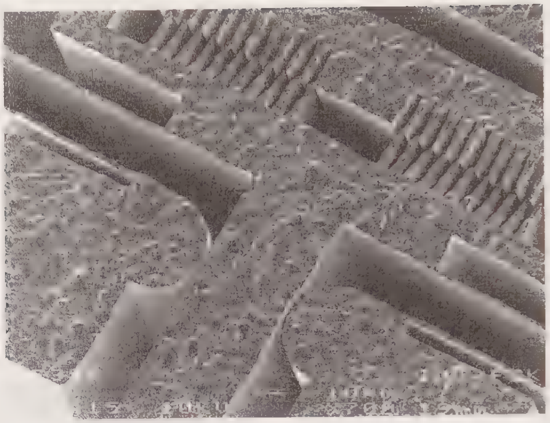


Fig. 5.35. Part of an electrostatic comb-drive-type linear actuator fabricated using electrodeposition of Ni. The PMMA resist has been removed.[86]

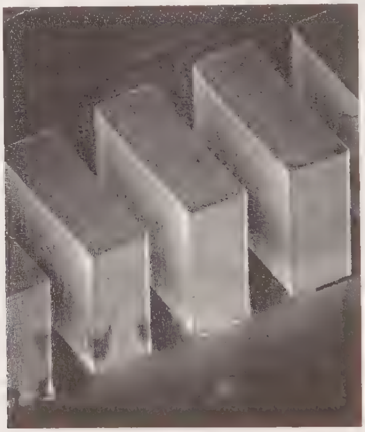


Fig. 5.36. SEM picture of Permalloy studs electroplated at IBM through PMMA resist to a height of 150 μm. Note the defects in the lower regions of the left hand side studs.

bottom region of the stud on the extreme left shows evidence of a defect, possibly related to incomplete removal of PMMA resist during the development step. The Permalloy was deposited using a solution typically used in thin film head manufacture[58,59]. At IBM, we have electrodeposited Permalloy (80at%Ni) up to thicknesses of 250μm, as shown in the SEM in Fig. 5.37 of part of a minimotor rotor plated in a PMMA resist stencil provided by MCNC and CAMD. On the left of this picture is shown the hole for the motor pin bearing; the other openings are used to facilitate solution release of the rotor from the substrate. Around the openings, the Permalloy is slightly thicker than the main body of the rotor because of current crowding.

In MEMS, copper is mainly used in devices requiring good electrical conductors. Displayed in Fig. 5.38 is an SEM of 60μm thick and 60μm wide Cu conductors electrodeposited through x-ray-exposed PMMA, which was then removed. As expected for such resists, the walls are vertical, and there is negligible underplating, indicating good adhesion of PMMA to the Ti/Cu/Ti seed layer. Copper lines with the same dimensions, electroplated through a positive resist exposed optically (Fig. 5.39) have walls with angles (ca. 82-85 degrees) significantly less than those observed in the x-ray exposed part. It is generally difficult to approach 90° wall angles with these optical positive resists. Although negative resists such as Shell Epon SU 8 are much better in this regard, they are difficult to remove following electrodeposition.

The results shown in Fig. 5.39 are very good for positive resist; we have also obtained resist patterns in excess of 100μm in thickness having similar wall angles. We have successfully combined the use of positive resist with X-ray exposed PMMA for multilevel structure fabrication (in association with CAMD and MCNC). In such hybrid cases, the PMMA resist process is used for tall structures, especially those in which gap tolerances are important.

Bischofberger et al.[88] from the Institute of Microsystems Technology in Buchs, and Allen et al. from the Georgia Institute of Technology[89] have investigated a high-aspect-ratio, polyimide-based negative resist that is *optically exposed* as a possible low-cost alternative to x-ray exposed PMMA resist for fabrication of small and medium sized structures. Using a spray procedure for developing the patterns, the former claim to have achieved structures with heights of 300μm, lateral dimensions of 30μm, and sharp sidewalls for aspect ratios up to 10. Figure 5.40 shows an example of a planar coil made using electrodeposited Cu. The Cu is 130 μm high, 90μm wide, and the air gap is 22μm.

The acid Cu and sulfamate Ni electroplating solutions employed by Bischofberger et al. are shown in Table 5.2. Details of the methods used for controlling the Ni solution were briefly discussed by the authors: pH control, monitoring of Ni concentration using either polarography or titration, and measurement of the sodium dodecyl sulfate surfactant concentration using a potentiometric tenside titration. These workers utilized

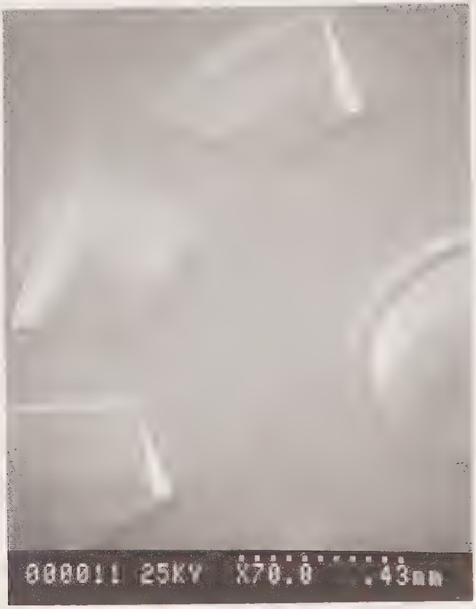


Fig. 5.37. SEM of part of a Permalloy minimotor rotor electrodeposited using x-ray exposed PMMA resist, which was then removed. The thickness of the Permalloy is in excess of 200 µm.

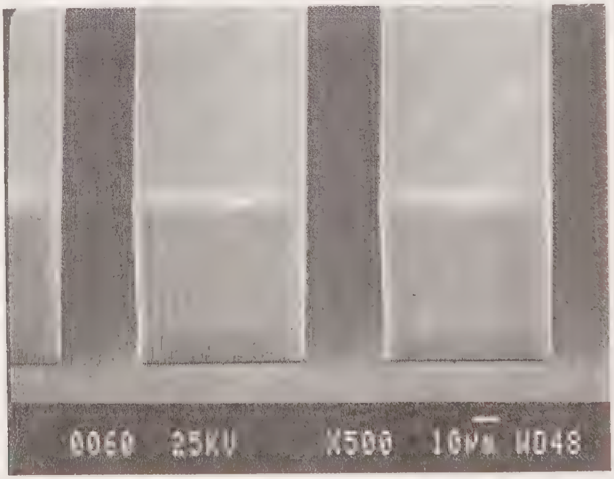


Fig. 5.38. SEM of minimotor Cu conductors, 70 µm thick, that were electrodeposited through x-ray exposed PMMA resist, which was then removed.

an acid Ni strike to cathodically clean the seed layer and to lay down a thin, strongly-adhering Ni layer in preparation for the main layer of Ni. As discussed earlier, Bischofberger et al. reported that a plasma cleaning

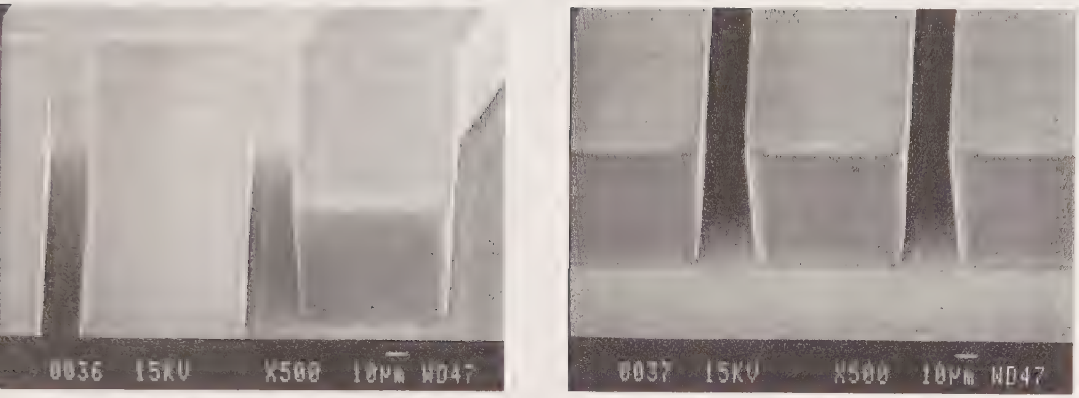


Fig. 5.39. SEM of minimotor Cu conductors, 58 µm thick, that were electrodeposited through Shipley SJR 3740 positive resist, which was then removed.

step was required after resist development prior to the electroplating step; they employed an O_2/SF_6 plasma for this.

Allen et al.'s work on photosensitive polyimide preceded Bischofberger et al.'s and was confined to thinner resist patterns that exhibited nearly vertical sidewalls in developed patterns. Figure 5.41 shows an assembled gear/pin structure made from electroplated Cu. The gear was plated on a different substrate from the pin and chemically released. Chromium was used as the sacrificial release layer. A notable feature of the assembled structure in Fig. 5.41 is the narrow gear/pin gap of less than 2µm.

Table 5.2: Compositions and operating conditions of the Cu and Ni electroplating solutions used by Bischofberger et al. [87]

	Cu(II)-sulfate pentahydrate		Ni(II)-sulfamate tetrahydrate	
Bath Temp.		25°C		50-60°C
Current Dens.		2.5 A/dm^2		5-6 A/dm^2
Bath Comp.	Cu sulfate	40-60 g/l	Ni(II)-sulfamate	95-110 g/l
	H_2SO_4	70-90 g/l	Cl	1.5 g/l
	Cl	40-120 mg/l	Boric acid	30-40 g/l
	Micrograin Cu73	5 g/l	Sodium dodecyl sulfate	4 g/l

Besides offering hope for a low-cost alternative to x-ray LIGA, the work of Bischofberger et al.[88], Allen et al.[89], and others illustrates that electrodeposition can be carried out using a variety of different resists in a high-aspect-ratio mode since it seems to be possible to be able to adequately clean the seed layer in fine patterns in the various resist materials before immersion in the plating bath.

5.8 FABRICATION OF MAGNETIC MINIMOTORS

In previous sections, the procedure of electroplating was discussed at some length, and fabrication processes involving single x-ray exposures, electroplating, and release or partial release of a single layer, were alluded to along with those of more complicated multilayer structures. A good example of the latter was the integrated magnetic microactuator of Ahn and Allen[85] (Figs. 5.33 and 5.34) which integrated electroplating with a variety of fabrication steps, e.g., involving polyimide dielectric and cantilever release. However, it is not enough to be able to electroplate tall structures and exhibit good uniformity of deposit thickness and composition control in the case of composition-sensitive alloys such as Permalloy. One must adhere to good process integration practice, espe-

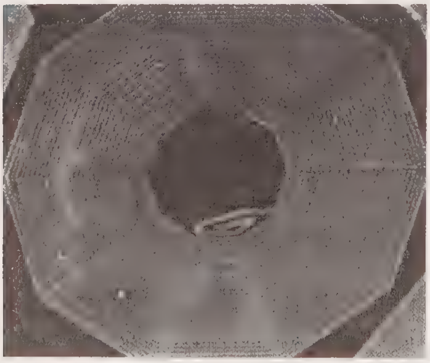


Fig. 5.40. A planar coil fabricated using Cu electroplating by Bischofberger et al.[88] Height = 130 µm; Cu width = 90 µm; air gap = 22 µm.

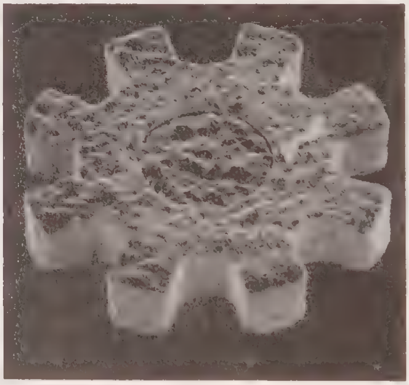


Fig. 5.41. Electroplated Cu gear/pin structure. The gear was formed on a different substrate, chemically released and placed onto the pin. The gear and pin heights are 50 µm, and the gear/pin gap is less than 2 µm.[89]

cially when fabricating structures with more than one level. Examples of phenomena to be aware of in integrating an electroplating process in a multilevel fabrication process include: (a) deposit stress, especially in the case of thick deposits, (b) seed layer cleanliness such that all key features plate successfully, especially vias, (c) good seed layer adhesion, which becomes important in the case of thick, or stressed deposits, (d) avoidance of overheating in the physical vapor deposition method, especally diode sputtering, used in seed layer deposition, which can cause stress in multilayer structures, (e) that chemical etchants used to remove seed layers or sacrificial layers do not attack the underlying dielectric in multilayer structures.

The authors have referred to the electroplating and fabrication efforts of several groups in this chapter; however, it has not been possible to give a detailed description of each group's work, especially regarding process integration, because of space limitations. Instead it was decided to provide descriptions, with an emphasis of process integration, of the fabrication details challenges encountered for two integrated, variable reluctance magnetic minimotors fabricated in their laboratory. One of these was built using high-aspect-ratio optical lithography for the stator, with x-ray lithography being used for the rotor; the other was built using x-ray lithography mainly, with optical lithography being used for minor levels.

5.8.1 Fabrication Overview — Optical

The motor, which is 17 mm in diameter, consists of six horseshoe-shaped magnets surrounding a rotor. Helical coils of Cu conductors are wound around the magnets.[90] Key physical parameters of the motor are given in Table 5.3.

TABLE 5.3
MINI-MOTOR PARAMETERS

Motor size diam. (mm)	Rotor radius (μm)	Core width (μm)	Coil width (μm)	Gap between coil (μm)	Air gap between rotor and stator (μm)	Coil turns (per core)
17	2976	700	60	30	5-20	108

The minimotor design[91] and magnetic field modeling[92] have been described elsewhere. The rotor and the stator were fabricated separately. The objective of this was to have a selection of rotors with different gap separations to not only build the motor but to check how well our predictive model held. The stator fabrication sequence is summarized in Fig. 5.42. Five photomasks were used. The motor consisted of the following levels:

1. Bottom-level Cu conductors;
2. Polymer insulation layer between bottom conductor - magnetic core;
3. Magnetic core, via, pin bearing;
4. Polymer insulation layer between magnetic core - top conductor;
5. Top-level Cu conductors.

The fabrication process for the stator was initiated by sputtering a Ti/Cu/Ti seed layer on an oxidized Si wafer, lithographically forming the plate-through mask for the bottom part of the Cu conductor windings, removing the upper Ti film, and electrodepositing Cu. Following removal of resist and seed layer, a dielectric was applied, and the surface was planarized, leaving the top surfaces of the plated Cu features exposed. A photo-patterned dielectric layer was then formed on top of the Cu conductors to insulate them from the magnetic core, leaving the end of each Cu bar exposed to permit coil formation in subsequent steps. A Ti/Cu/Ti seed layer was then deposited; photoresist was applied and patterned; the upper Ti film was chemically etched away; and the magnetic horseshoe core, rotor pin bearing, and coil vias were fabricated in one Permalloy electrodeposition step. The resist and seed layer were removed; dielectric was applied, cured, and polished back to the top of the Permalloy features, leaving a planar surface. A second photo-patterned dielectric layer was formed over the core regions, leaving the vias exposed. Once again, seed layer was deposited; resist was applied and patterned; and the top Cu conductor level was electrodeposited to complete the conductor coils. The structure could then be encapsulated by a final dielectric application, if desired. The rotor was fabricated separately.

5.8.2 Lithography — Optical

High-aspect-ratio optical or x-ray lithography was employed to form the various resist patterns, the latter for the rotor pattern. In the course of this work we sought to determine how far optical resists could be extended in terms of thickness and aspect ratio.

The positive working resist, a high-viscosity Novolac obtained from Shipley (SJR 3740), or Hoechst (AZ P4620), was applied using a multilayer spin method to achieve thicknesses up to 75μm. Hexamethyl disiloxane (HMDS) was used as an adhesion promoter to the underlying Ti/Cu/Ti seed layer. A slow baking procedure with good temperature control was necessary to stabilize the resist prior to UV exposure and to prevent stress-related defects. Resist patterning was achieved using contact printing with a standard (manual) mask aligner (Carl Suss). After a single exposure, a commercially available, alkaline-type developer from

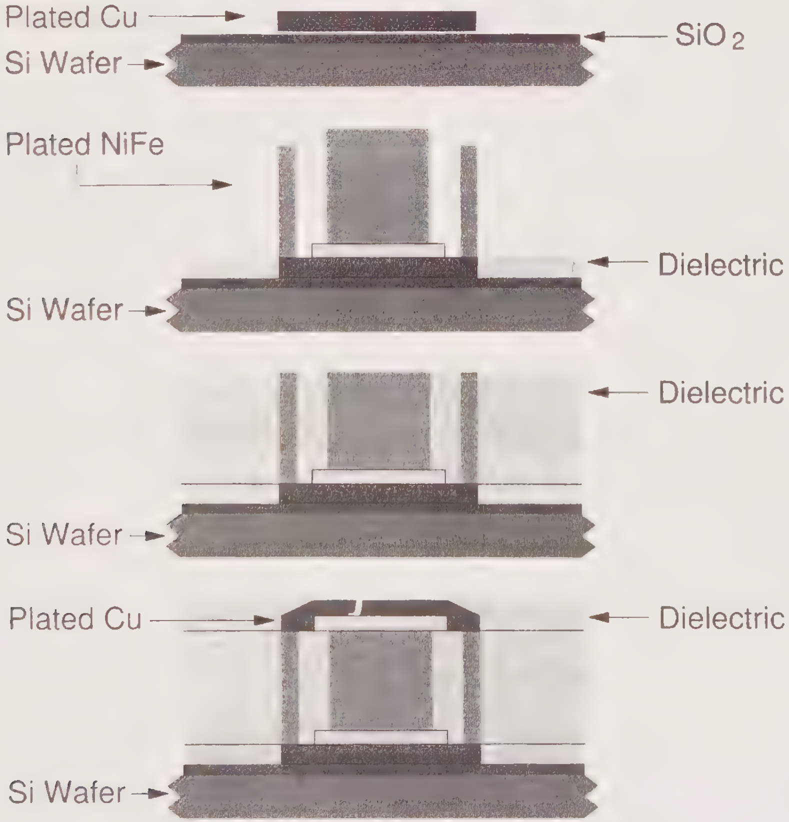


Fig. 5.42. Schematic representation of the fabrication flow for the minimotor stator. Seed layers are not shown for clarity.

Shipley (Microposit 2401 - CD 275) was used to develop the pattern features. Straight side walls of about 80° were obtained.

5.8.3 Permalloy Electrodeposition

The soft magnetic material in the horseshoe-shaped cores and rotors was Permalloy (Ni(80 at%)/Fe). Deposition was carried out in a horizontal paddle cell, with the paddle between electrodes travelling at a distance of about 3 mm above the cathode surface, and a frequency of 1 Hz. The plating cell was connected to a larger plating solution reservoir, and the solution was continuously recirculated and filtered using 0.2μm Millipore filters. The temperature in the reservoir was maintained at 270.1°C. Solution pH was controlled at 30.1 through automatic addition of dilute HCl solution as needed. Using a Brinkmann Dosimat, the ferrous ion concentration was kept constant by adding a pH=2 solution of ferrous sulfate at a rate calculated to match the rate of iron consumption. A nickel plate served as the anode. A stainless-steel plate - generally coated with a plated layer of Permalloy - surrounding the cathode served as an auxiliary electrode, or "thief".

The current densities used were in the 5-12.5 mA/cm^2 range, based on the active area of the wafer. Auxiliary electrode current densities were set at 1.1 - 1.2 times the superficial, or geometric, device wafer current densities, thus ensuring uniform deposit thickness across the wafer. Individual features, however, did exhibit higher thickness at the Permalloy-resist boundary, caused by "current crowding", especially when features, or areas of features, were surrounded by a large area of resist, as in the case of the stator pin-bearing. This behavior is consistent with the earlier discussion of feature-scale uniformity. In general, Permalloy deposits were smooth and faithfully reproduced the lithographic patterns.

5.8.4 Copper Electrodeposition

Copper was deposited in a paddle cell using a commercially available acid-Cu plating solution operated at the conditions recommended by the vendor. At the target current density, the deposition rate was 20 μm/hr. An annulus around the periphery of the wafer, left bare of photoresist, served as an on-wafer "thief electrode". Although this on-wafer thief was not powered independently, good thickness uniformity was achieved;

across-wafer thickness variation was typically about 3%. The thickness uniformity depended primarily on the layout of the photoresist pattern. Since the current efficiency of Cu deposition was 100%, the deposit thickness was easily controlled once the active area of the photoresist pattern was accurately known.

5.8.5 Dielectric Materials and Planarization — Optical

Various polymer systems, including polyimide, hard-baked photoresist and epoxy resins, were investigated as dielectric fillers and as interlevel insulation layers for the minimotor. An epoxy resin-based dielectric system was successfully used, which required much lower temperatures for complete curing, was solvent-free, and eliminated the void/delamination problem. Planarization was carried out after each application and curing of the dielectric.

Fabrication of the interlevel insulation layer was an important step in the motor fabrication process since it isolates the magnetic core from the bottom and top Cu conductors. This was accomplished by spin-coating a thin layer of epoxy, 5-7μm in thickness, on the planarized substrate. This was then patterned, leaving behind a band of epoxy (Fig. 5.43) which either covered all but the ends of the Cu conductors for via connections in the case of the bottom coil conductors, or covered the exposed, polished, Permalloy core leaving the vias bare in the case of the middle, magnetic core level of the motor.

5.8.6 Completion of the Optical Minimotor Structure

The rotor cavity in the stator was opened by reactive ion etching the epoxy dielectric with masking to ensure selectivity; dielectric removal rates were approx. 0.5μm per minute. Figure 5.44 shows an SEM of portion of a stator after removal of the epoxy. Removal of the latter was cleanly achieved in the rotor region, although some epoxy removal also occurred in the region of the top-level coils due to plasma access under the mask.

Sacrificial Cu release layers were employed to facilitate rotor release. A thin seed layer was sputtered on the oxidized Si wafer, and a blanket Cu film 1.0 μm thick was electroplated. The photolithography and rotor Permalloy plating were completed on the Cu-coated substrate. Rotor re-

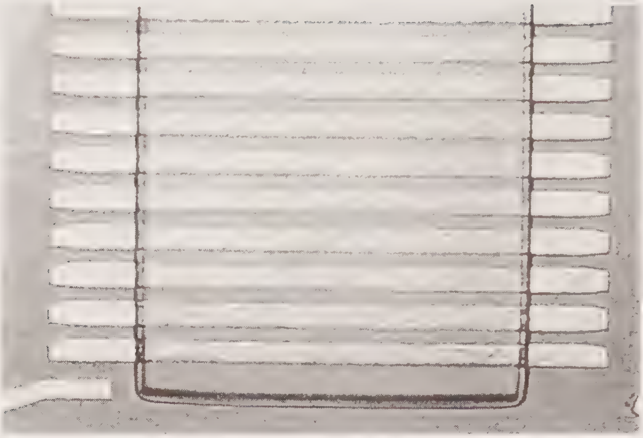


Fig. 5.43. Bottom minimotor coil structure after application of the patterned insulation layer used to insulate the Cu conductors for the Permalloy cores.

lease was accomplished by selective dissolution of the Cu in a near-neutral, ammonium persulfate solution.

The completed minimotor (partial view of assembled motor in Fig. 5.45) was operated in an open-loop configuration, with maximum rpm of (880) which was limited by the undamped resonance of the rotor. The rotor levitation threshold was 60 mA. The motor torque, as determined from the angular dependence of inductance, was 2.3 x 10-7 N-m at 100 mA 0-pk for a 43μm thick Permalloy core. A more detailed description of the operation of the motor is available elsewhere.[94]

5.8.7. X-Ray Minimotor Fabrication

The objectives were to design a high-torque, magnetic minimotor and to develop a batch fabrication process to build a motor using LIGA, x-ray lithography, and electroplating. The previously developed lumped parameter model[91], 3D magnetostatic modeling[92], and test results and fabrication experience gained on the optical were used to design the present motor. The fabrication technology and the choice of materials had to be modified considerably to deal with problems associated with the x-ray lithography approach.

Figure 5.46 illustrates the construction of the motor, which basically consists of a bottom Cu level, a permalloy core level, a top copper level,

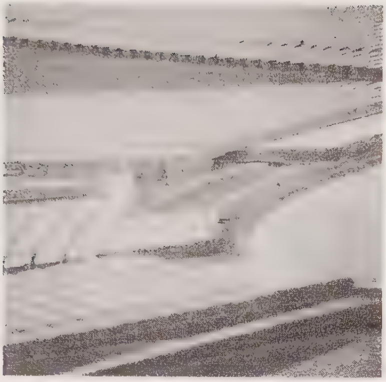


Fig. 5.44. SEM of portion of a completed stator after removal of the dielectric from the rotor region using RIE.

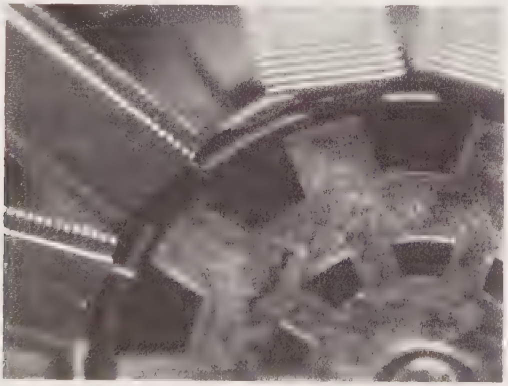


Fig. 5.45. Optical picture of a part of the optical-lithography-fabricated IBM minimotor with rotor in place.

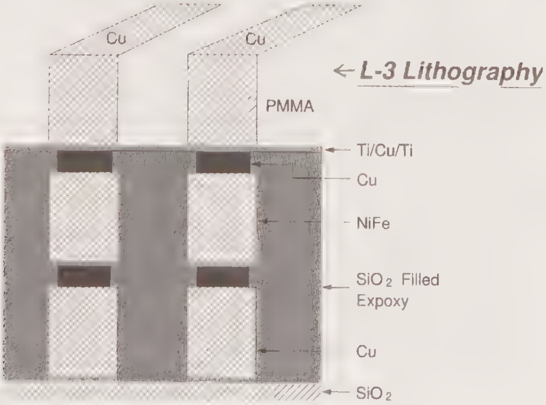

Cu
Cu
← L-3 Lithography
PMMA
Ti/Cu/Ti
Cu
NiFe
SiO2 Filled Expoxy
Cu
SiO2

Fig. 5.46. Schematic representation of the cross-section of a partially completed x-ray lithography fabricated minimotor stator.

0 0326 25kV X110 27μm

Fig. 5.47. Bottom Cu conductor patterns with 20 μm thick Cu separator pads after removal of the AZ and PMMA resists and the seed layer. Application of the Si-filled epoxy is the next processing step.

and insulation layers to isolate the Cu conductors from the core. Each of the metal levels (a Cu level is shown in detail in Fig. 5.47) was built by applying a seed layer, spinning on and curing 2μm or less of PMMA, gluing a PMMA sheet (Perspex CQ), fly-cutting to thin down the 1 mm sheet down to the desired thickness, x-ray exposure, development, pretreatment of the surface, plating Cu or NiFe, and planarizing. The process was continued by applying AZ optical resist, UV exposing, developing, plating thin metal pads, removing the AZ, PMMA and seed layers, backfilling the structure with a dielectric, curing and planarizing the surface in preparation for the seed layer, and x-ray lithography steps of the next level.

The above sequence was carried out three times to complete the bottom Cu, permalloy, and top Cu levels. Vias connecting the top and the bottom Cu conductors were made of NiFe plated at the same time as the magnetic core. The rotor shaft was plate up in stages each time that a stator feature was plated. X-ray lithography was used to form the Cu coils and to achieve the demanding wall profiles on the rotor and stator. The simpler optical resist process provided adequate definition for the thin metal pads which bridged the insulation levels between the copper winding and the core. The rotors were plated on a separate substrate through x-ray exposed PMMA, detached and slipped over the central shaft of the stator. The top view of the stator is shown in Fig. 5.48.

5.8.7.1 Process Compatibility With X-Ray Lithography

The use of x-ray lithography, which allows significantly thicker structures to be built, poses certain challenges:

1. The thicker layers exacerbate thermal mismatch problems between the copper coils, permalloy yoke, and the dielectric. The thermal mismatch in the Z direction (perpendicular to the substrate) becomes a serious issue. A perfectly planarized layer containing metal features embedded in a polymer, usually with high TCE, can suffer delamination at the vertical metal-polymer interface when heated to 180°C during the cure of the spun-on PMMA, and can also cause cracking and delamination of the seed layer. This problem was overcome by using a dielectric filled with low TCE inorganic.
2. Using a dielectric containing inorganic particles to reduce TCE necessitated changes in the RIE process used to open the rotor cavity in the stator.
3. The 100 to 400μm thick PMMA created alignment problems because of the difficulty of accurately registering a mask to alignment marks under the thick resist. Techniques for alignment through up to 400μm thick PMMA with +/-5μm accuracy were developed in a joint effort between MCNC, CAMD and IBM. The alignment entailed developing a jig to permit optical alignment, clamping of the x-ray

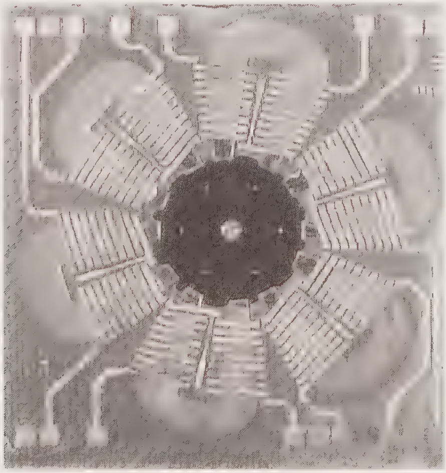


Fig. 5.48. A top view of a completed x-ray-fabricated minimotor stator (rotor not in place).

mask to the substrate, and forming windows in the resist to be able to see the specially designed alignment marks without distortion.

4. Successful plating of small features in the thick PMMA required special attention to the entire sequence of steps from the development of the PMMA through the plating operation. A high yield in plating, particularly of the 100x140μm studs through the thick PMMA mask was achieved by developing a sequence of: surface pretreatment, PMMA development, and wetting initiation steps that assured uniform onset of plating in all the resist openings at the same time.
5. All solvents and chemicals used in PMMA processing had to be compatible with materials incorporated in the underlying structure.
6. Successful lamination of the PMMA sheet as described in the fabrication section above required that irregularities in the planarized wafer surface be minimized. Producing the required degree of planarization necessitated development of appropriate polishing techniques to deal with the various composite surfaces at each level.

The motor described in this work represents the first time that a series of mutually compatible process steps, including three aligned x-ray exposures, have been integrated into a total process to produce a working, five-level, magnetic MEMS motor.

5.9 DISCUSSION

Electrodeposited Cu and Ni have generally been the metallic materials of choice in MEMS, the latter often for magnetic as well as structural uses, since it is a hard material. More recently, Permalloy (Ni(80wt%)/Fe) has been replacing Ni as the magnetic material of choice because of its higher saturation magnetization (10K Gauss for 80:20 NiFe compared to 6K Gauss for Ni). Electrodeposition of these materials has been described earlier and will not be discussed here.

5.9.1 Minimum Requirements To Be Able to Plate Through the Mask or Practice LIGA

The authors and their colleagues have learned through many years of experience in developing and applying plating-through mask technology to a variety of experimental structures and manufactured products that the successful use of this technology requires careful attention to seven critical areas. These are:

- *Seed Layer* — its interaction with resist, substrate, developer, and plating solution.
- *Resist Layer* — its relation and interaction with the seed layer, developer, plating bath, etc.
- *Mask Layout* — some of its requirements for good thickness uniformity.
- *Electroplating Solution* — requirements for the plating solution and its interaction with seed layer, resist, etc.
- *Plating Tool* — some key requirements to achieve good results.
- *Resist Removal Process* — resist stripper interaction with the metallurgy and the dielectric.
- *Seed Layer Removal Process* — its interaction and effect on the pattern generated by plating and integrity of the device.

The requirements in each area are stated in Table 5.4. Many of these rules may seem obvious to one who has practiced plating through mask technology for some time. However, the table should provide invaluable guidance for those new in the field, and even experienced worker may find it useful to review this table, especially when a process is not working as expected.

Table 5.4 Effect of various process variables on the ability to generate metal patterns by different patterning techniques. What conditions must be satisfied?

Seed Layer - Must:

1. Have good adhesion to the substrate.
2. Have high electrical conductivity.
3. Not be readily oxidized or contaminated.
4. Have ability to withstand immersion in the developer and in plating solution without excessive etching or oxidation.
5. Be easily selectively removable when plating is completed.
6. Not produce ionic species which poison the plating solution.
7. Not react with resist but must provide good adhesion for it.

Resist Layer - Must

1. Be pore free when applied.
2. Be easily selectively exposed by UV, E-Beam, X-Ray or other radiation means.
3. Be readily penetrated by radiation.
4. Show high dissolution (etch) ratio by developer between the exposed and unexposed regions.
5. Produce upon development nearly vertical walls.
6. Not leave residue on the seed layer in developed areas.

7. No react with the plating solution.
8. Not release etchables which can contaminate the plating solution.
9. Show good adhesion to the seed layer.
10. Not be under excessive stress (must not crack, peel, etc.)
11. Be easy to remove when plating is completed.

Pattern - Must

1. Be spaced uniformly.
2. Be not excessively varying in size and/or shape.

Electroplating Solution - Must

1. Be easy to plate from.
2. Be capable of producing good, quality metal (electrical, magnetic, resistance, etc.)
3. Not attack the resist, substrate or seed layer.
4. Readily wet in to smallest recesses produced in resist (not leave entrapped air bubbles).
5. Not produce excessive amount of H_2 during plating.
6. Have a good throwing power.
7. Be capable, with appropriate tools, of producing uniform thickness and composition of deposit.
8. Not be very sensitive to local current density variations.
9. Be easy to replenish and control over long periods of time.

Plating Tools - Must:

1. Provide good current density distribution.
2. Provide uniform agitation.
3. Be well protected from particulate and chemical contamination.
4. Include filtration, temperature, pH, and analytical controls.
5. Include provisions for adding fresh reactants.

Resist Removal Process - Must:

1. Not result in chemical attack of the plated pattern or the seed layer.
2. Be easy to remove cleanly without leaving residue.

Seed Layer Removal Process - Must:

1. Be selective - not attack the plated pattern.
2. Remove both conducting and adhesion films.
3. Not leave residue or surface contamination.

Space does not permit discussion of each of the issues in Table 5.4. However, some discussion will be included on two highly critical issues. Even with this discussion, the individual reader will have to consider the stated requirements and make his own observations in the context of the specific system being used because issues will vary slightly from application to application. The best guides will be the reader's own experience and the extensive literature which is being created on plating-through-mask applications.

5.9.2 Resist Residue

Any residue left after resist development will affect the film nucleation and adhesion; wetting, especially in very deep, round vias; H_2 gas entrapment; and uniformity of current distribution. A few of these issues are discussed here.

With regard to wetting, it should be noted that it is always harder to fill circular or square vias with electrolyte than it is to fill elongated slots. Elongated slots and interconnected patters will, once wetting starts, fill by capillary action. With good development of the resist and with adequate rinsing, wetting is usually not a problem even for very tall structures.

One interesting observation which we can offer with regard to residue: when residue is too thick (>30 to 40Å), nucleation will not start. The best remedy is one more rinse in clean developer and/or oxygen plasma ash.

5.9.3. Plating into Deep Recesses — Current and Mass Distribution

The other important issue that one has to consider is mask design, pattern density, density distribution, and the ratio of the height (depth)-to-width of the features. Figures 5.49(a) and 5.49(b) show schematically the differences when plating through thin or thick mask. In thin masks, because of current crowding the large areas will grow with a raised rim near the interface with the resist. The current cannot spread out quickly enough so due to higher current density near the resist interface there is higher rate of plating. In the case of small vias the current crowding will result in excessive growth of the vias. For a normal rate of agitation the merging of rims into raised areas will occur when the opening is below 5μm in diameter. The rate of growth will be slower for long lines (two directional current crowding) vs. round vias which see a three directional current crowding.

In cases where resist is thick by comparison to the dimensions of the pattern, the situation is different. During the initial plating in a >1:1 ratio depth-to-width pattern, the current crowding is not as pronounced, apparently because of a certain degree of current spreading in the via during

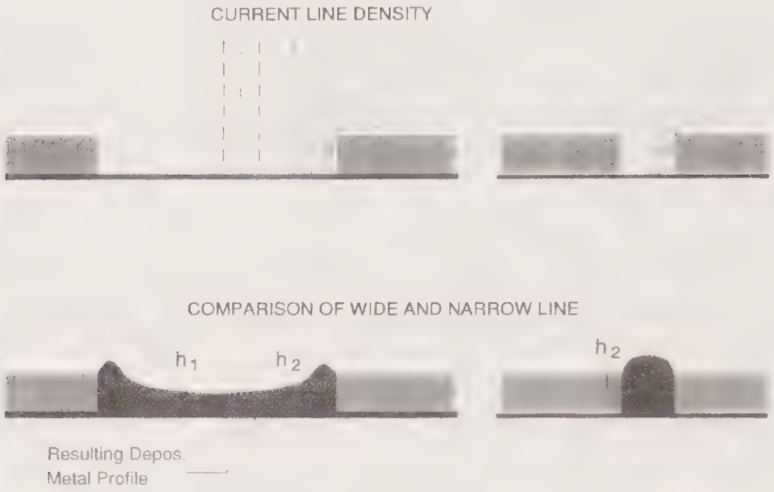

CURRENT LINE DENSITY
COMPARISON OF WIDE AND NARROW LINE
h_1
h_2
h_2
Resulting Depos.
Metal Profile

Fig. 5.49(a). This schematic representation shows the current lines that are expected to be obtained during electrodeposition through a thin resist pattern.

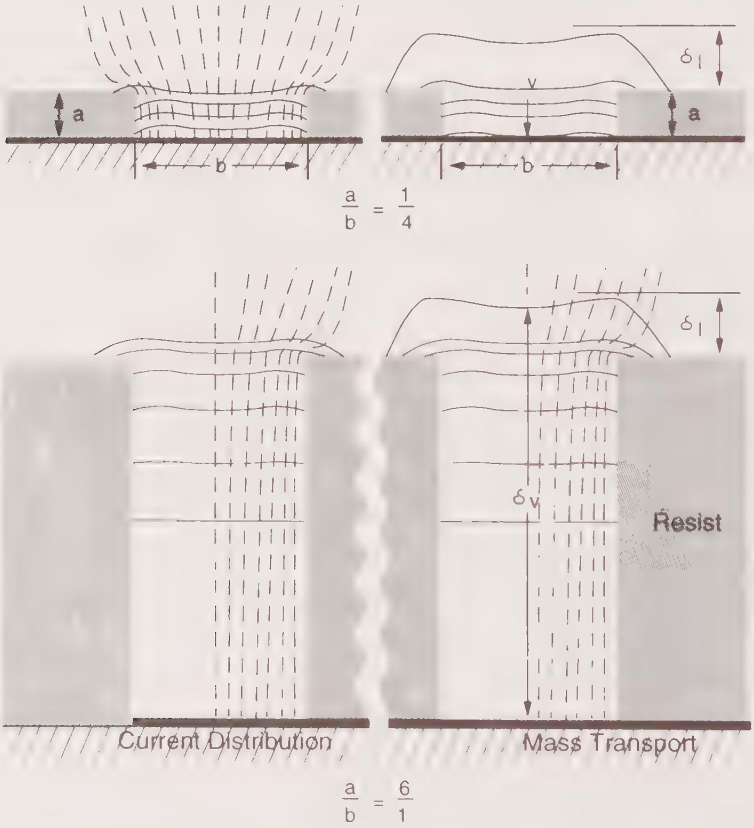

a
b
a
b
δ_l
v
$\frac{a}{b} = \frac{1}{4}$
δ_l
δ_v
Resist
Current Distribution
Mass Transport
$\frac{a}{b} = \frac{6}{1}$

Fig. 5.49(b). This figure shows the current lines that are expected to be obtained during electrodeposition through a thick resist pattern.

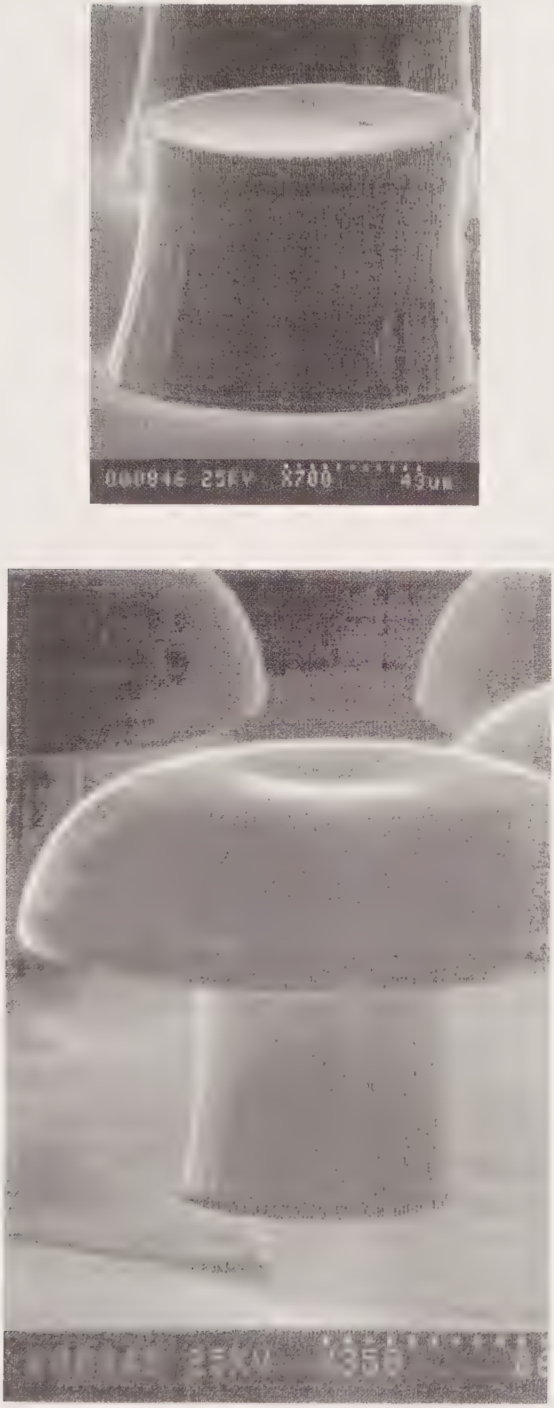

000946 25KV X700 43um
25KV X350

Fig. 5.50. (a) Effect of current crowding at a feature edge on thickness uniformity. Diameter of feature is approximately 100 μm. (b) Deposit type obtained as result of severely overplating the feature in (a). The resist has been removed in both cases.

the plating of the first ~2/3 of the via where only a small rim is observed. This is shown in Fig. 5.50(a). When the Permalloy film continues to grow, the depth-to-width ratio decreases, and current crowding near the edges becomes pronounced. As growth continues a mushroom shape is formed. The shape shown in Fig. 5.50(b) reflects the path of the current flux lines. Since they are perpendicular to the surface, a mushroom head is formed. All around the edge there is higher current whereas in the center of the mushroom head there is much less current crowding, hence the dimple in the mushroom head.

These kinds of problems can be minimized by increasing the concentration of the current suppressant in the electroplating solution, and by operating in the secondary current distribution region. When plating in the primary current distribution region, or near the tertiary, the situation is even more aggravated.

Therefore, if one deals with plating into deep recesses, it is advisable to stop plating at ~ 2/3 of the total depth. Also it is preferred to have as small differences between the feature sizes as possible. It is also advisable to design the mask such that the patterns are as uniformly distributed as possible. If it is desired to build a next layer on top of the first one, it become necessary to planarize the structure after plating and before resist removal. Plating can be stopped at any time if planarization is to follow, although it is preferred to stop at the approximate height that is actually desired.

While plating-through-masks in connection with electronic devices is now relatively well understood and is easy to control, the same cannot be said of the plating into very deep recesses such as is often required for MEMS devices. The successful fabrication of one device or even 10 wafers does not constitute a controlled manufacturing process. The LIGA or plating through very thick masks is possible and can be refined but it will require considerable support work in chemical bath formulation, equipment design and clean development of the resists from the recesses. In general, one can state that if the masks can be designed in such a way that the resist area will be minimized and the plated area maximized, there is virtually no problem with plating. If the situation is reversed, and one will have to mask very large areas and clean residue out of very small diameter holes, considerable care and precautions will have to be taken to achieve good yield.

5.10 SUMMARY

The plating through mask technology has made tremendous advances in the last three decades. It has become a very important method of electronic device fabrication. When it was necessary to increase the bit density in the electronic computer components, the electroplating through mask came to the rescue by providing the ability to grow very thick deposits necessary to form conductors and devices. Because of the ability to achieve the 3D structures it is possible to batch fabricate low-cost inductive magnetic thin film heads. Plating through mask technology helped x-ray lithographic masks become reality, which in turn helped demonstrate the ability to fabricate 3D metal structure with height-to-width aspect ratios greater than two orders of magnitude. Plating through mask technology after three decades in the electronic research laboratories and more than one decade in the industrial products in the form of the thin film magnetic heads today is finding its way into MEMS under the name of LIGA.

The plating through mask technology, "LIGA", has become a very important enabling technology which when combined with surface and bulk silicon micromachining gives immense new possibilities to those working in the field of MEMS. It makes possible electroplating of copper, gold, silver, and also of magnetic metals and metal alloys such as Ni, NiFe, CoNi, etc. Therefore micromechanical structures whose principle of operation is based not only on electrostatic, elastic, or other silicon based properties, but in addition on properties provided by magnetic metals and alloys, can be used. This opens up, to the researchers in the field of micromechanics, a much larger arsenal of materials, properties, and techniques to build microsensors and actuators.

It was the objective of this chapter to review the historical aspects of developments of the plating through mask technology in electronics, the experience gained by the researchers working in the electronic industry in the field of magnetics and electrochemistry, and in particular the plating through mask technology, so that the researchers in the field of MEMS could benefit from the knowledge, findings, and ideas accumulated in connection with plating of electronic structures.

As a result of the electronic industry initiative, a considerable amount of work was done in the past three decades both by the industry and by the industry supported studies at universities. These studies provide a better understanding of the plating-through-mask technology and the knowledge transformed in many situations electroplating from the state of science. Mathematical models have been developed that explain a variety of phenomena and provide guidelines for cell design, preferred pattern layout on the mask, and predict the shape evolution of the patterns. Still, much remains to be done to make the plating-through-mask technology rigorously scientific and completely predictable in every situation. In other cases, experience, intuition, and common sense are still the most useful allies when a quick solution is to be achieved while exploring a new system.

Suggested Further Reading:

1. J. O'M. Bockris and A.K.N. Reddy, "Modern Electrochemistry", Vols. **1** and **2**, Plenum Press, New York, (1972).
2. F.A. Lowenheim, "Modern Electroplating", John Wiley & Sons, New York, (1974).
3. L.J. Durney, "Electrochemical Engineering Handbook", Fourth Edition, Van Nostrand Reinhold Co., Inc., New York, (1984).
4. E. Budevski, G. Staikov, W.J. Lorenz, "Electrochemical Phase Formation and Growth - Introduction to the Initial Stages of Metal Deposition", VCH Book Series, New York, (1996).
5. P.C. Andricacos and L.T. Romankiw, "Magnetically Soft Materials in Data Storage: Their Properties and Electrochemistry", in Advances in Electrochemical Science and Engineering, Edited by H. Gerischer and C.W. Tobias, **3**, 230-312 (1994).
6. L.T. Romankiw and D.A. Thompson, "Magnetic Properties of Plated Films" in Properties of Deposits, Their Measurement and Significance, The Electrochemical Society, Inc., **23**, 389-427 (1976).
7. L.T. Romankiw, "Pattern Generation in Metal Films Using Wet Chemical Techniques: A Review", in Etching for Pattern Definition, Pub. by The Electrochemical Society, Inc., Pennington, New Jersey, pp. 161-93, (1976).
8. The Electrochemistry Society (ECS) proc. volume (PV) ECS PV-87-17, "Electrodeposition Theory and Practice", by L.T. Romankiw and D. Turner.

9. ECS PV-88-23, "Electrochemical Technology in Electronics", L.T. Romankiw and T. Osaka.
10. ECS PV-90-8, "Magnetic Materials Processes and Devices", L.T. Romankiw and D.A. Herman, Jr.
11. ECS PV-92, "Magnetic Materials, Processes and Devices", L.T. Romankiw and D.A. Herman, Jr.
12. ECS PV-94-6, "Magnetic Materials, Processes and Devices", L.T. Romankiw and D.A. Herman, Jr.
13. ECS PV 95-18, "Magnetic Materials, Processes and Devices", L.T. Romankiw and D.A. Herman, Jr.

Acknowledgments

The authors wish to acknowledge the contributions of and numerous discussions with more than 60 collaborators with whom they had the privilege to interact and work during the past many years (35 years in the case of L.T.R.). Without their contributions, active discussions and suggestions much less would have been accomplished, and the IBM Center for Electrochemical Technology and Microfabrication in Yorktown Heights, NY would not have achieved the level of knowledge, experience and expertise which it now has. The number of collaborators goes far beyond what is possible to list in this acknowledgment. Many of the names appear in the references cited, although even that is far from being complete. The authors would also like to acknowledge the support from HI-MEMS ARPA Alliance, which sponsored in part the work on the minimotor using synchrotron radiation and x-ray lithography. The authors acknowledge with deep appreciation the invaluable help of M.P. Barringer in preparing this manuscript. They also acknowledge with thanks the contributions of S. Krongelb in reviewing several sections of this chapter.

REFERENCES

1. L.T. Romankiw, "Evolution of the Plating Through Lithographic Mask Technology", ECS proc. **95-18**, Edited by L.T. Romankiw and Dean Herman, Jr., Pennington, New Jersey, pp. 253-273, (1995).
2. L.T. Romankiw, "Electroforming for Electronic Devices", Plating and Surface Finishing, **84**(1), 10-16 (1997).
3. L.T. Romankiw, "Pattern Generation in Metal Films Using Wet Chemical Techniques: A Review", in Etching for Pattern Definition, Pub. by The Electrochemical Society, Inc., Pennington, New Jersey, pp. 161-93, (1976).
4. J. O'M. Bockris and A.K.N. Reddy, "Modern Electrochemistry", Vols. **1** and **2**, Plenum Press, New York, (1972).
5. F.A. Lowenheim, "Modern Electroplating", John Wiley & Sons, New York, (1974).
6. L.J. Durney, "Electrochemical Engineering Handbook", Fourth Edition, Van Nostrand Reinhold Co., Inc., New York, (1984).
7. E. Budevski, G. Staikov, W.J. Lorenz, "Electrochemical Phase Formation and Growth - Introduction to the Initial Stages of Metal Deposition", VCH Book Series, New York, (1996).
8. H. Dahms and I.M. Croll, "The Anomalous Codeposition of Iron-Nickel Alloys", J. Electrochem. Soc. **112**, 771 (1965).
9. P.C. Andricacos and L.T. Romankiw, "Magnetically Soft Materials in Data Storage: Their Properties and Electrochemistry", in Advances in Electrochemical Science and Engineering, Edited by H. Gerischer and C.W. Tobias, **3**, 230-312 (1994).
10. L.T. Romankiw and D.A. Thompson, "Magnetic Properties of Plated Films" in Properties of Deposits, Their Measurement and Significance, The Electrochemical Society, Inc., **23**, 389-427 (1976).
11. H. Deligianni and L.T. Romankiw, "In Situ pH Measurement During Electrolysis Using a Rotating pH Electrode", IBM J. of Res. and Devel., **37**, 85-95 (1993).
12. S. Hessami and C.W. Tobias, "A Mathematical Model for Anomalous Codeposition of Nickel-Iron on a Rotating Disk Electrode", J. Electrochem. Soc., **136**, 771 (1990).

13. K.Y. Sasaki and Jan B. Talbot, "Modelling Studies of the Electrodeposition of the Binary Iron-Group Alloys", Proc. ECS Magnetic Materials, Processes and Devices, Application to Storage and MEM's, Edited by L.T. Romankiw and D.A. Herman, Jr., **95-18**, 685-716 (1995).

14. K-M. Yin, B.N. Papov, and R.E. White, "Electrodeposition Modeling of Nickel-Iron Alloys in Presence of Organic Additives", Proc. AESF International Conference, Atlanta, GA, June 22-25, p. 103, (1992).

15. W. Reidel, "Electroless Nickel Plating", ASM International, Metals Park, Ohio (1991).

16. G.O. Mallory and J.B. Hajdu, eds., Electroless Plating: Fundamentals and Applications, American Electroplaters and Surface Finishers Society, Orlando, Florida (1990).

17. E.K. Yung, L.T. Romankiw, R. Alkire, "Plating into Through Holes and Blind Holes", J. Electrochem. Soc., **136**(1), 206-15 (1989).

18. M. Paunovic, "Electroless Deposition of Metals and Alloys", in Electrochemistry in Transition, Edited by Oliver J. Murphy and B. Conway, Plenum Press, New York, pp. 479-97 (1992).

19. P. Bindra, S.L. Levine, W.T. Pimbley, R.R. Schaffer, J. Reid, W.L. Underkofler, R.T. Galasco, "Metal Deposition", in Principles of Electronic Packaging, Edited by D.P. Seraphim, R. Lasky, C-Y. Li, McGraw Hill, Inc., New York, Ch. 16, 470-518 (1989).

20. V.R. Markovich, F.S. Poch, "Plated Through-Hole Technology", Ch. 17, pp. 519-538 in ref. 19.

21. O.A. Moreno, R.C. McHatton, R.S. Paonessa, R.H. Magnuson, "Etching of Metals by Wet Processes in Electronic Packaging", Ch. 18, pp. 539-76 in ref. 19.

22. R.R. Tummala, R.W. Keyes, W.D. Grobman, S. Kapur, "Thin Film Packaging", in Microelectronics Packaging Handbook, Edited by R.R. Tummala and E.J. Rymaszewski, Van Nostrand Reinhold Co., New York, Ch. 9, pp. 673-725 (1989).

23. D.P. Seraphim, D.E. Barr, W.T. Chen, G.P. Schmitt, R.R. Tummala, "Printed-Circuit Board Packaging", Ch. 12, pp. 853-921 in ref. 22.

24. J. Kim, S.H. Wen, D.Y. Jung, R.W. Johnson, "Microstructure Evolution During Electroless Copper Deposition", IBM J. Res. and Dev., **24**(6), 697-710 (1984).

25. G. Ganu and S. Mahapatra, "Electroless Gold Deposition for Electronic Industry", in J. of Sci. & Ind. Res., **46**, 154-161 (1987).

26. H. Ali and I. Christie, "A Review of Electroless Gold Deposition Processes", in Gold Bull., **17**, 118-127 (1984).

27. Y. Okinaka and C. Wolowodiuk, "Electroless Gold Deposition: Replenishment of Bath Constituents," in Plating, **58**, 1080-1084 (1971).

28. D.F. Canaperi, R. Jagannathan, M. Krishnan, "Method of Replenishing Electroless Gold Plating Bath," US Patent 5,635,253 (1997).

29. J.O'M. Bockris and G.A. Razumney, Fundamental Aspects of Electrocrystallization, Plenum Press, New York (1967).

30. W.H. Safranek, "Properties of Electrodeposited metals and Alloys", American Electroplaters and Surface Finishers Society, Orlando, Florida (1986).

31. D. Pletcher, "Industrial Electrochemistry", Published by Chapman and Hall, Ltd, New York (1984).

32. S. Rahskov, D.S. Stoichev and I. Tomov, Electrochim. Acta, **17**, 1955 (1972).

33. K. Ragunathan and R. Weil, Surf. Technol., **10**, 33 (1980).

34. R. Weil, "The Origins of Stress in Electrodeposits: Review of Literature Dealing with Stress", American Electroplaters Society, Inc., Research Report 60, Feb. (1971).

35. A.Y. Vagramyan and Z.A. Soloveva, Technology of Electrodeposition, Robert Draper Ltd, Teddington, England (1961).

36. M. Pourbaix, Lectures on Electrochemical Corrosion, Plenum Press, New York (1973).

37. D.A. Jones, Principles and Prevention of Corrosion, Macmillan, New York (1992).

38. H.H. Uhlig and R.W. Revie, Corrosion and Corrosion Control, John Wiley & Sons, New York (1985).

39. V.I. Lainer, Modern Electroplating, Translated from Russian to English, published by the Israel Program for Scientific Translations, available from US Dept. of Commerce (1970).

40. H.L. Pinkerton, "Current and Metal Distribution", in Electroplating Engineering Handbook, Ed. by L.J. Durney, 4th Edition, Van Nostrand Reinhold, New York (1984).

41. J. Hyner, "Design for Plating, in Electroplating Engineering Handbook, 4th edition, Ed. by L.J. Durney, Van Vostrand Reinhold, New York (1984).
42. S. Mehdizadeh, J. Dukovic, P.C. Andricacos and L.T. Romankiw, and H.Y. Cheh, "Optimization of Electrodeposit Uniformity by the Use of Auxiliary Electrodes", J. Electrochem. Soc., **137**, 110 (1990).
43. J.V. Powers and L.T. Romankiw, "Electroplating Cell Including Means to Agitate the Electrolyte in Laminar Flow", US Patent 3,652,442 (1972).
44. L.T. Romankiw, PC. Andricacos, J.O. Dukovic, "Multi-Compartment Electroplating System", US Patent 5,312,532, (1994).
45. S. Mehdizadeh, J. Dukovic, P.C. Andricacos, L.T. Romankiw and H.Y. Cheh, "The Influence of Lithographic Patterning on Current Distribution: A Model for Microfabrication by Electrodeposition", J. Electrochem. Soc., **139** 78, (1990).
46. S. Mehdizadeh, J. Dukovic, P.C. Andricacos, L.T. Romankiw and H.Y. Chen, "The Influence of Lithographic Patterning on Current Distribution in Electrochemical Microfabrication", J. Electrochem. Soc., **140**, 3497 (1993).
47. L.T. Romankiw, "Electrochemical Technology in Electronics Today and its Future: A Review", Oberfläche-Surface, pp. 238-47 (1984).
48. L.T. Romankiw, "Electrochemical Technology in Electronic Industry and It's Future", New Materials and Processes, JEC Press, Inc., **3**, 39-51 (1985).
49. L.T. Romankiw, "A Path from Electroplating Through Lithographic Masks in Electronics to LIGA in MEMS", Electrochim. Acta, (1997), in press.
50. L.T. Romankiw, I.M. Croll and M. Hatzakis, "Batch Fabricated Thin Film Magnetic Recording Head", IEEE Trans. Mag-6, No. 4, p. 729 (Sept. 1970).
51. L.T. Romankiw, "Thin Film Inductive Heads: From One to Thirty-One Turns", ECS Proc. Symp. on Magnetic Materials, Processes and Devices, Ed. L.T. Romankiw and D.A. Herman, Jr., PV **90-8**, 39-53 (1990).
52. L.T. Romankiw, S. Krongelb, E.E. Castellani, A.T. Pfeiffer, B. Stoeber and J.D. Olsen, "Advantages and Special Considerations in

Fabricating Bubble Circuits by Electroplating and Sputter Etching", IEEE Mag-10, No. 3, pp. 828-31, Sept. 1974.

53. M.C. Blakeslee, R.E. Acosta, S. Krongelb, L.T. Romankiw and B.J. Stoeber, "Electrodeposition Process for Fabrication of the Conductor in First SLM 2μm Bubble Memory", J. Electrochem. Soc., **125**(3), 152 (1978).

54. L.T. Romankiw, E.E. Castellani, and M. Hatzakis, "Additive Electroplating Limits for Gold Conductors Using PMMA Electron Beam Resist", Extended Abstracts, No. 208, Spring Mtg. of The Electrochem. Soc., San Francisco, CA, May (1974).

55. E. Spiller, R. Fader, J. Topalian, E.E. Castellani, L.T. Romankiw and M. Heritage, "X-Ray Lithography for Bubble Devices", Solid State Technology, pp. 62-78, April 1976.

56. R. Feder, E. Spiller and J. Topelian, "Replication of 0.1μm Geometries with X-Ray Lithography", J. Vac. Sci. Techn., **12**(6), 1332 (1975).

57. L.T. Romankiw, "Eliminating of Undercut in an Anodically Active Metal During Chemical Etching", US Pat. 3,853,715 (1974).

58. N.C. Anderson and C.R. Grover, Jr., "Electroplating of Nickel-Iron Alloys for Uniformity of Nickel/Iron Ratio Using a Low Density Plating Current", US Patent 4,289,707, July 21 (1981).

59. E.E. Castellani, J. Powers, L.T. Romankiw, "Nickel-Iron (80:20) Alloy Thin Film Electroplating Method, Electrochemical Treatment and Plating Apparatus", US 4,102,756, July 25 (1978).

60. L.T. Romankiw, "Use of A-Z Type Photoresist as Insulation in Multi-Turn Thin Film Heads", IBM Technical Disclosure Bulletin, Nov. 1980.

61. J.V. Powers, L.T. Romankiw, "Electroplating Cell Including Means to Agitate the Electrolyte in Laminar Flow", US Pat. 3,652,442 (1972).

62. E.E. Castellani, I.M. Croll, A.T. Pfeiffer and L.T. Romankiw, "Fine-Line Circuit Fabrication and Photoresist Application Therefore", US Pat. 4,315,985 (1982).

63. R.E. Fontana, Jr., "Magnetic Thin Film Heads, A Review on Processing Issues", ECS Proc., Ed. by L.T. Romankiw and Dean A. Herman, Jr., PV **90-8**, 205 (1990).

64. M.E. Bowden and J.K. Smith, "American Chemical Enterprise: A Perspective on 100 Years of Innovation to Commemorate the Centennial of the Society of Chemical Industry — The Information Age and the Environment — Revolution in Electronics", pp. 79-80 (1994)

65. K. Ohashi, M. Ito, M. Watanabe, "Application of Electroplating to Thin Film Heads: An Overview", ECS Proc., Ed. by L.T. Romankiw and T. Osaka, PV **88-23**, 525-42 (1988).

66. A.M. Tuxford and L.T. Romankiw, "A Feasibility Study of Planar Multilevel Structure Using Electrodeposited Gold and Organic Dielectrics", ECS Extended Abstracts, No. 223, Fall Meeting (1974).

67. S. Krongelb, J.O. Dukovic, M.L. Komsa, S. Mehdizadeh, L.T. Romankiw, P.C. Andricacos, A.T. Pfeiffer and K. Wong, "The Application of Electrodeposition Process to Advanced Fabrication", SPIE Proc. Microelectronic Interconnects and Packages: Optical and Electrical Technologies, **1389**, 249 (1990).

68. M. Datta, R.V. Shenoy, C. Jahnes, P.C. Andricacos, J. Horkans, J.O. Dukovic, L.T. Romankiw, J. Roeder, H. Deligianni, H. Nye, B. Agarwalla, H.M. Tong and P. Totta, "Electrochemically Robust PbSn C4 Interconnects", J.ECS, **142**(11), 5779-83 (1995).

69. LIGA News, Editor W. Ehrfeld, First Issue, January (1995).

70. E.W. Becker, W. Ehrfeld, P. Hagmann, A. Maner, D. Müchmeyer, "Fabrication of Microstructure with High Aspect Ratios and Great Structural Heights by Synchrotron Radiation Lithography", Galvanoforming and Plastic Moulding, LIGA Press.

71. W. Ehrfeld, P. Bley, F. Götz, P. Hagmann, A. Maner, J. Mohr, H.O. Moser, D. Münchmeyer, W. Schelb, D. Schmidt, E.W. Becker, "Fabrication of Microstructures Using the LIGA Process", IEEE (1987).

72. R. Ruprecht, W. Bcher, J. Hauselt, V. Pioter, "Injection Molding of LIGA and LIGA-Simular Micro-Structures Using Filled and Unfilled Thermoplastics", proceedings Micromachining and Microfabrication Process Technology, SPIE **2639** 146-57 (1995).

73. R. Ruprecht, K. Müller, J. Konys, V. Piotter, A. Theis, H. Wöllmer, J. Hnsslet, "Various Replication Techniques for Manufcturing Three-Deimensional Metal Microstructures", HARMST-97 Conf., June 20-21, Madison, WI (1997).

74. A. Ruzzu, J. Farenberg, M. Heckele, T. Schaller, "Multifunctional Valve Components Fabricated by Combination of LIGA Processes and Mechanical Microengineering", MARMST-97 Conf., June 20-21, Madison, WI (1997).

75. H. Schift and J. Söchtig, "LIGA Technology for the Fabrication of Positioned Planar Structures", HARMST-97 Conf., June 20-21, Madison, WI (1997).

76. U. Wallrabe, P. Bley, B. Krevet, W. Menz, J. Mohr, "Theoretical and Experimental Results of an Electrostatic Micro-Motor with Large Gear Ratio Fabricated by LIGA Process", IEEE Micro-Electro-Mechanical Systems '92, Travemünde, Germany, Feb. 4-7 (1992), IEEE Control No. 92CH3093-2.

77. H. Guckel, K.J. Skrobis, T.R. Christenson, J. Klein, S. Han, B. Choi, E.G. Lovell and T.W. Chapman, "Fabrication and Testing of the Planar Magnetic Micromotor", J. Micromech. Microeng. **1**, 135-38 (1991).

78. A. Rogner, W. Ehrfeld, D. Münchmeyer, P. Bley, C. Burbaum, and J. Mohr, "LIGA-based flexible microstructures for fiber-chip coupling", J. Micromech. Microeng. **1**, 167-70 (1991).

79. M. Harmening, W. Bacher, P. Bley, A. El-Kholi, H. Kalb, B. Kowanz, W. Menz, A. Michel, J. Mohr, "Molding of Three Dimensional Microstructures by LIGA Process", Micro-Electro-Mechanical Systems '92, Travenmünde, Germany, Feb. 4-7 (1992).

80. D.R. Sparks, S.R. Zarabadi, J.D. Johnson, Q. Jiang, M. Chia, O. Lassen, W. Higdon and P. Castillo-Borelley, "A CMOS Integrated Surface Micromachines Angular Sensor: It's Automotive Applications", IEEE Transducers '97 International Conf. on Solid State Sensors and Actuators 3B1.04, p. 851-54 (1997).

81. Douglas R. Sparks, "Electroformed Micromachines Automotive Applications", MST NEWS VDI/VDE Technologiezentrum Informationstechnik GmbH, Rheinstrasse 10B, D14513 Teltow, MST NEWS 20/87, p. 4 (1997).

82. T. Ohnstein, Z. Cook, H. Guckel and J. Klein, "Demonstration of of a Micro-Electro-Mechanical Tunable Filter for Infrared Spectroscopy", HARMST-97, Madison, Wisconsin, June 20 (1997).

83. B. Chaudhari, H. Buckel, J. Klein, K. Fisher, "Photoresist Application for the LIGA Process", to be published in the Research J. of Microsystems Technology (1997).

84. J. Klein, H. Guckel, B. Chaudhari, K. Fisher, Y. Ahn, "An Asynchronous Micro-Induction Motor", HARMST-97, Madison, WI, June 20-21 (1997).

85. C.H. Ahn and M.G. Allen, "A Fully Integrated Surface Micromachined Magnetic Microactuator with a Multilevel Meander Magnetic Core", J. Micromechan. Syst., **2**(1), 15 (1993).

86. A. Maciossek, B. Lochel, H.-J. Quenzer, B. Wagner, S. Schulze and J. Noetzel, "Galvanoplating and Sacrificial Layers for Surface Micromachining", Microelectron. Eng., **2**(2), 503 (1995).

87. M.W. Borner, M. Pohl, F.J. Patenburg, W. Bacher, H. Hein, and W.K. Schomberg, Microelectronic Engineering **30** 505, (1995).

88. R. Bischofberger, H. Zimmermann and G. Staufert, "Low-Cost HARMS-Process", Extended Abstract EUROSENSORS '96, Leuven; also preprint entitled "NTB Low-Cost LIGA-Process", with authors R. Bischofberger and G. Staufert.

89. A.B. Frazier and G.M. Allen, J. Micromechanical Systems **2**(2), 87 (1993).

90. R.E. Acosta, C. Ahn, I.V. Babich, T.J. Chainer, E.I. Cooper, J.M. Cotte, S. Hegde, W.J. Horkans, C. Jahnes, S. Krongelb, K.T. Kwietniak, N.C. LaBianca, E.J. O'Sullivan, A.T. Pomerene, D.L. Rath, L.T. Romankiw and J.A. Tornello, "Integrated Variable Reluctance Magnetic Minimotor", Proc. of Fourth Intern. Symp. on Materials, Processes, and Devices, Edit.: L.T. Romankiw and D.A. Herman, Jr., The Electrochemical Society, Pennington N.J. PV **95-18**, 508-520, (1995)

91. T.J. Chainer, L.T. Romankiw, "An Integrated Variable Reluctance Magnetic Motor", ECS Proceedings, PV95-18, pp. 482-92 (1995), Edited by L.T. Romankiw and D. A. Herman, Jr. (1995).

92. P.L. Trouilloud, "Magnetostatic 3D Modeling of a Magnetic Micromotor", in ref. 90, pp.493-507 (1995).

93. L.T. Romankiw, S. Krongelb, E.E. Castellani, J. Powers, A.T. Pfeiffer and B. Stoeber, "Additive Electroplating Technique for Fabrication of Magnetic Devices", Proc. Int'l. Conf. on Magnetics ICM-73, Vol. VI, pp. 104-11 (1973), Moscow, USSR.

94. M. Hatzakis, "Electron Resists for Microcircuit and Mask Production", J.ECS, Vol. 116, pp. 1133-37 (1969).

95. N.C. LaBianca, J.D. Gelorme, E.I. Cooper, E.J. O'Sullivan and J. Shaw, in ECS Proc. Fourth Intl. Symp. on Magnetic Materials, Processes and Devices (1995), L.T. Romankiw and D. Herman Editors, PV **95-18** (1996) 386.

96. B. Loechel, A. Maciossek, H.J. Quenzer, B. Wagner, "Electrochemical Microfabrication II", Proc. ECS PV 94-32, p. 100, Edited by M. Datta, K. Sheppard, J. Dukovic. (1994).

97. T.J. Chainer, L.T. Romankiw, "An Integrated Variable Reluctance Magnetic Motor", ECS Proceedings, PV95-18, pp. 482-92 (1995), Edited by L.T. Romankiw and D. A. Herman, Jr. (1995).

98. I.V. Babich, T. Chainer, E.I. Cooper, S. Hegde, J. Horkans, C. Jahnes, S. Krongelb, K.T. Kwietniak, N.C. LaBianca, E.J.M. O'Sullivan, L.T. Romankiw, J.A. Tornello, P. Trouilloud, "Variable Reluctance Magnetic Integrated Mini-Motor: Fabrication and Test Results", INTERMAG-97, to be published in IEEE Trans Mag. 1997.

CHAPTER 6

High Aspect Ratio Processing

Craig R. Friedrich
Robert Warrington
Michigan Technological University

Walter Bacher*
Werner Bauer**
Philip J. Coane†
Jost Göttert†
Thomas Hanemann
Jürgen Haußelt**
Mathias Heckele
Regina Knitter**
Jürgen Mohr*
Volker Piotter
Hans-Joachim Ritzhaupt-Kleissl**
Robert Ruprecht

**Institut für Mikrostrukturtechnik (IMT)*
*** Institut für Materialforschung III (IMF III)*
Forschungszentrum Karlsruhe GmbH

†Louisiana Tech University

CONTENTS

6.1 LIGA Process / 301
    6.1.1 Introduction / 301
    6.1.2 Deep-etch x-ray lithography / 303
        6.1.2.1 Principle of operation / 303
    6.1.3 Mask technology for deep-etch x-ray lithography / 304
        6.1.3.1 Absorber / 304
        6.1.3.2 Membranes / 307
    6.1.4 Processes for deep-etch x-ray lithography / 309
        6.1.4.1 Fabrication of thick resist layer / 309
        6.1.4.2 Radiation chemistry and resist development / 311

        - 6.1.4.3 Exposure parameters for PMMA resists / 312
        - 6.1.4.4 Accuracy limits / 315
        - 6.1.4.5 Fresnel diffraction and photo electrons / 316
        - 6.1.4.6 Divergence of synchrotron radiation / 317
        - 6.1.4.7 Tilt of absorber side walls / 318
        - 6.1.4.8 Mask distortion / 318
        - 6.1.4.9 Fluorescence radiation from the mask membrane / 318
        - 6.1.4.10 Secondary electrons produced in the substrate layer / 319
    - 6.1.5 Modified LIGA processes / 321
        - 6.1.5.1 Sacrificial layer technique with the LIGA process / 321
        - 6.1.5.2 3D patterning in deep x-ray lithography / 323

- 6.2 Micromechanical Machining Processes / 325
    - 6.2.1 Introduction / 325
    - 6.2.2 Diamond machining / 328
    - 6.2.3 Microdrilling / 333
    - 6.2.4 Micromilling / 338

- 6.3 Plastic Molding Processes for Manufacturing LIGA and LIGA-Like Microstructures / 345
    - 6.3.1 Introduction / 345
    - 6.3.2 Molding microstructures from plastics / 346
        - 6.3.2.1 Filling the mold / 346
        - 6.3.2.2 Releasing the mold / 347
        - 6.3.2.3 Cycle time / 347
        - 6.3.2.4 Residual layer / 347
        - 6.3.2.5 Materials requirements / 348
    - 6.3.3 Tools and machines for plastic molding of microstructures / 348
        - 6.3.3.1 Mold inserts for plastic molding of microstructures / 348
        - 6.3.3.2 Injection molding / 350
        - 6.3.3.3 Hot embossing / 351
        - 6.3.3.4 Reaction molding / 352
    - 6.3.4 Molded microparts / 353
        - 6.3.4.1 Microfluidics / 354
        - 6.3.4.2 Micro-optics / 354
        - 6.3.4.3 Sensors and actuators / 357
        - 6.3.4.4 Lost micromolds / 359

- 6.4 Development of Ceramic Microcomponents / 360
    - 6.4.1 Introduction / 360
    - 6.4.2 Thermal two-stage process for the production of multicomponent ceramic powders / 361
    - 6.4.3 Shaping techniques / 361
    - 6.4.4 Development of microcomponents / 364
    - 6.4.5 Conclusions / 365

- 6.5 Summary / 365
    - References / 369

The development of silicon microcomponents and microsystems, being an outgrowth of the processing techniques commonly used for the fabrication of integrated circuits, was constrained to mostly planar dimensions. This was because optical lithography and most photoresists were developed for this specific purpose. In many applications, however, the components require that the vertical dimensions be larger than those obtained by traditional optical lithography methods. This led to the development of high aspect ratio processing in which the vertical dimension can be large compared to the lateral dimensions. While there is no apparent agreed-upon definition of high aspect ratio, the processes are generally characterized by x-ray lithography and/or micromechanical machining methods, with postprocessing and replication methods suitable for the dimensions involved. While there have been many accomplishments toward making these processes as economically viable as integrated circuit processes, much work remains to be done. The following sections describe the high aspect ratio processes and replication techniques. The work is a compilation of efforts worldwide and is meant to provide an overview of the processes and techniques.

6.1 LIGA Process

6.1.1 Introduction

The LIGA process was initially developed in the late 1970s at Forschungszentrum Karlsruhe (KfK, now FZK) as an alternative way to produce very small slotted nozzles for uranium isotope separation [1,2]. These separation nozzles needed to be of an arbitary shape with the smallest dimensions in the micrometer range. A large number of them had to be added in a cascaded array to achieve a high yield. The successful achievement of these requirements provided the impetus to apply the LIGA process to the fabrication needs of a broad range of processes in fields as diverse as communication technologies, automobile and medical technologies, or sensors for environmental, biological, and chemical analysis [3].

The LIGA process is schematically presented in Fig. 6.1. The acronym LIGA originates from the German expressions for the major process steps: deep-etch x-ray lithography with synchrotron radiation (LI), electroforming (G, Galvanoformung), and molding (A, Abformung).

In the first step the absorber pattern on an x-ray mask is transferred into a thick resist layer by shadow printing with synchrotron radiation. Forcing highly relativistic charged particles on a circular orbit leads to the emission of radiation. This situation occurs e.g. in electron storage rings when the electrons pass through the strong magnetic field of a dipole (bending) magnet. The intense radiation (so-called synchrotron radiation) is emitted tangentially to the electron orbit in the bending magnet and covers a wide continuum from the infrared to the x-ray region.

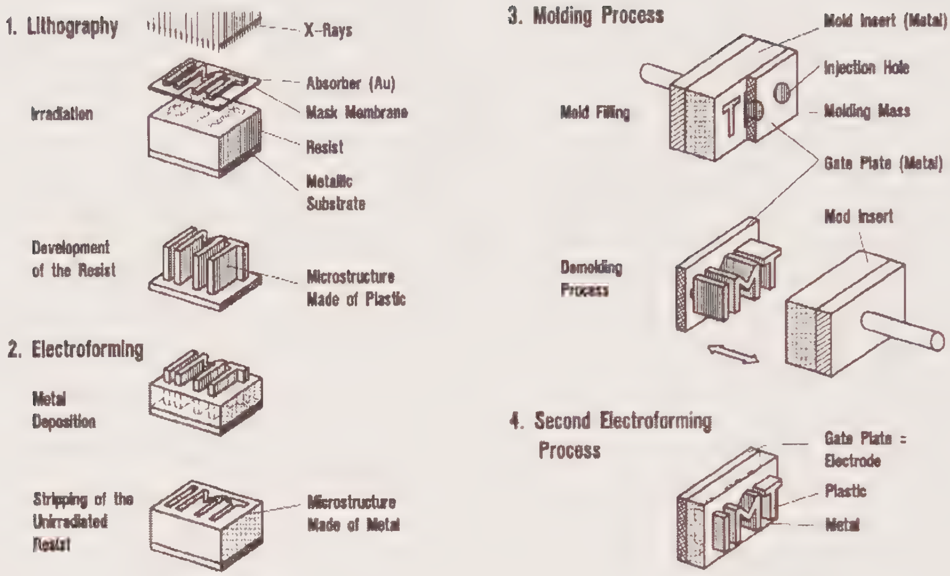

1. Lithography
X-Rays
Absorber (Au)
Irradiation
Mask Membrane
Resist
Metallic Substrate
Development of the Resist
Microstructure Made of Plastic
2. Electroforming
Metal Deposition
Stripping of the Unirradiated Resist
Microstructure Made of Metal
3. Molding Process
Mold Insert (Metal)
Injection Hole
Mold Filling
Molding Mass
Gate Plate (Metal)
Mold Insert
Demolding Process
4. Second Electroforming Process
Gate Plate = Electrode
Plastic
Metal

FIG. 6.1 Basic steps of the LIGA process.

The radiation is emitted into a narrow band with a vertical (vertical to the plane of motion of the electron) emission angle of typically 0.5 mrad. The radiated power and the spectral distribution can easily be calculated from the electron energy, the radius of curvature, the magnetic field, and the electron current in the ring. The thick resist layer is applied either on a massive metal plate or an isolating plate covered with a plating layer. Although the most commonly used x-ray resist is polymethyl methacrylate (PMMA), a chemically amplified negative resist that exhibits a much higher sensitivity has been developed and is under testing [4]. Depending on the resist thickness, the effective synchrotron spectrum used for pattern transfer has a characteristic wavelength ranging from 2 nm to 6 nm.

Within the irradiated sections, the x-ray photons are absorbed and initiate a material modification. In case of PMMA, the long molecular chains undergo main-chain scission and the solubility in this region is enhanced. By using a highly selective developer, these sections are completely dissolved, leaving the unirradiated areas, which form the primary pattern, completely intact. A highly selective developer is one where the dissolution rate difference between the exposed and unexposed resist is at least three orders of magnitude. By means of electroplating metals such as Cu, Ni, or Au, the complementary microstructure pattern is plated up using the conductive layer of the substrate as a cathode. Following this, the remaining PMMA is dissolved, leaving the patterned metal structures as the end product.

Alternatively, the electroplating may be extended to reach a thickness of 4–5 mm

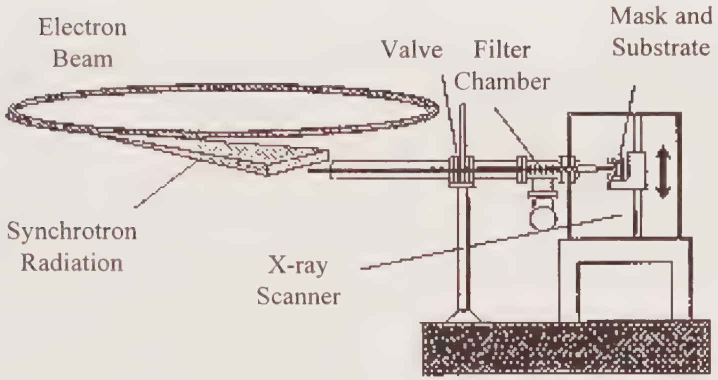


FIG. 6.2 Schematic of mask and substrate scanner, x-ray transport beamline, and synchrotron (not to scale).

above the height of the primary structure, producing a robust mold insert. This insert is replicated by molding techniques, such as reaction injection molding, injection molding, or hot embossing into polymers, offering nearly the same quality but at relatively lower cost and in much higher quantities compared with the primary pattern. These secondary polymer structures also have multiple applications. They may form the final polymer component, be used as a template in a second electroforming process, or serve as disposable molds for the fabrication of ceramic microstructures.

6.1.2 Deep-Etch X-Ray Lithography

6.1.2.1 Principle of operation

The schematic setup is shown in Fig. 6.2. An ultra-high vacuum beamline (x-ray transport line) is attached to the synchrotron ring. At the end, an exposure station is attached to the beamline. The exposure station is a vacuum chamber and houses a linear stage that moves the mask/substrate holder vertically and at a constant velocity through the narrow synchrotron beam. By using this scanning method, a homogeneous deposition of dose across the mask is ensured. Multiaxis scanning mechanisms are available that include tilt and rotate options and enable exposures in planes other than the perpendicular [5].

Typically, the ultra-high vacuum beamline is terminated by a beryllium window that separates the ring vacuum system from the vacuum system of the exposure station. A filter chamber located in front of the scanning mechanism accommodates different filters, which can be inserted into the beam in order to adapt the synchrotron spectrum to the resist thickness.

The mask and substrate are mounted on the motion stage in a special holder. Typically the mask is locked to a mask fixture, which is clamped to the stage. The proximity gap between the mask and the resist surface is defined by selecting an appropriate shim thickness. The substrate is pressed against the shims and locked into this position. The back side of the substrate fixture, as well as the mask fixture, is water-cooled to compensate for thermally induced dimensional changes resulting from the different materials used. Additional cooling of the mask membrane is achieved with a 100-mbar helium atmosphere in the exposure chamber during the exposure processs. Details about exposure stations, mask/substrate holders, and exposure conditions are presented in Refs. 5–10.

Beamlines as described, or slightly modified, are found attached to synchrotron sources all over the world [11]. Although they may differ in specific details, they have many common aspects such as the filter chamber, vacuum chamber, a single or multi-axis scanning mechanism, and a mask/substrate holder.

6.1.3 Mask Technology for Deep-Etch X-Ray Lithography

In deep-etch x-ray lithography (DXRL), x-ray masks are of major importance because they define the quality of the patterned microstructure. Unlike the masks used for soft x-ray lithography in complementary metal oxide semiconductor (CMOS) circuit fabrication, deep-etch x-ray masks require high-aspect ratio absorber structures to provide sufficient contrast between the mask membrane and absorber materials. Suitable membrane materials offer high transparency to x-rays and consist of low Z (atomic number) x-ray absorption materials such as beryllium, carbon (either graphite or diamond), silicon, or titanium; or compounds such as silicon nitride, silicon carbide, and boron nitride. On the other hand, the absorber pattern has to be made from high Z materials such as tungsten, gold, or tantalum, which have a high absorption coefficient for x-rays in the energy range from 1–2 keV up to 30 keV. The most commonly used material is gold because it can be easily electroplated and forms a stress-free film. This material provides a high x-ray absorption coefficient for a minimum structural height of several micrometers (5–50 μm). The synchrotron spectrum and the resist thickness will determine what height is required to ensure good exposed contrast.

6.1.3.1 Absorber

Depending on the structural details, there are different ways to generate an intermediate mask with a 3-μm-thick resist layer. Since the requirements for the smallest feature size are less stringent, the resist can be patterned by conventional optical lithography [this includes ultraviolet (UV) lithography] using an electron beam-written chromium mask. The tuning of the baking process as well as the exposure and development parameters have been shown to achieve nearly vertical side walls, 2-μm minimum features, and slightly rounded corners [12–13]. Although electroplated gold structures up to 10 μm high have been patterned, the

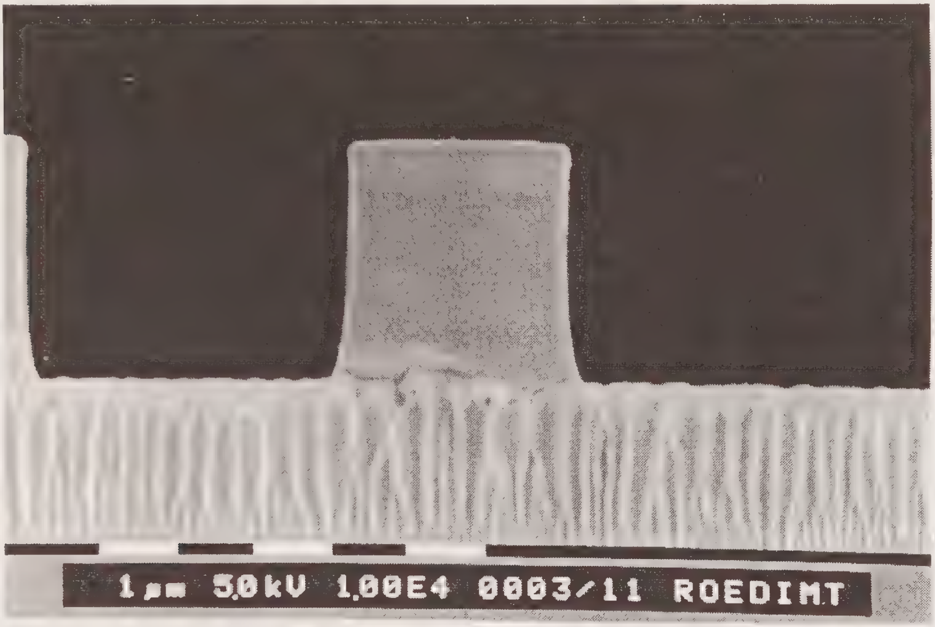


FIG. 6.3 Resist patterned by optical lithography technique for x-ray mask fabrication. Resist height is 3–3.5 µm and smallest feature size is 2.5 µm.

minimum possible feature size then increases to 5 µm [10]. Resist patterning with this technique is shown in Fig. 6.3.

A more precise way to pattern 3-µm-thick resist is by direct writing using a high-energy (100 keV) electron beam writer. Due to the high acceleration voltage, the broadening of the focused electron beam (proximity effect) occurs in the membrane layer, leading to nearly ideal vertical side walls in the resist pattern [14]. The highest accuracy in patterning is achieved with PMMA resist. However, the low sensitivity of this resist requires long exposure and extended writing times. In order to generate submicrometer structures, different layers are required to pattern the boundary (fine layer, small spot size) and the coarse areas (coarse layer, large spot size). A different dose must be deposited in each layer. The geometrical limits of this technology are lines and spaces of 1 µm width and a minimum feature size of 0.1–0.2 µm. Patterned resist using high-energy electron beam writing is shown in Fig. 6.4.

Another precise way is a trilevel process using reactive ion etching (RIE), with typical results shown in Fig. 6.5. Starting with a thin photoresist layer (100 nm), a high-precision pattern is generated either by optical or electron beam lithography. Next, a titanium layer (approximately 15 nm thick) is patterned by etching in an Ar plasma. Finally, a thick polyimide layer (3 µm) is patterned in an oxygen plasma. Minimum features of 0.3–0.4 µm have been achieved with this process. [15].

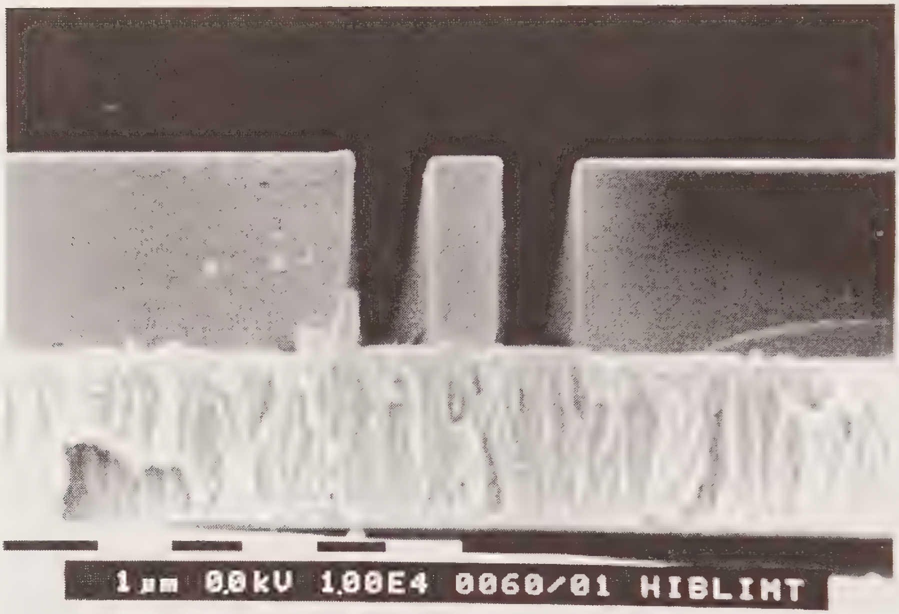

1 µm 0.0kV 1.00E4 0060/01 HIBLIMT

FIG. 6.4 Patterned resist using high-energy electron beam writing. Resist height is 3–3.5 µm and smallest feature size is 1.0 µm.

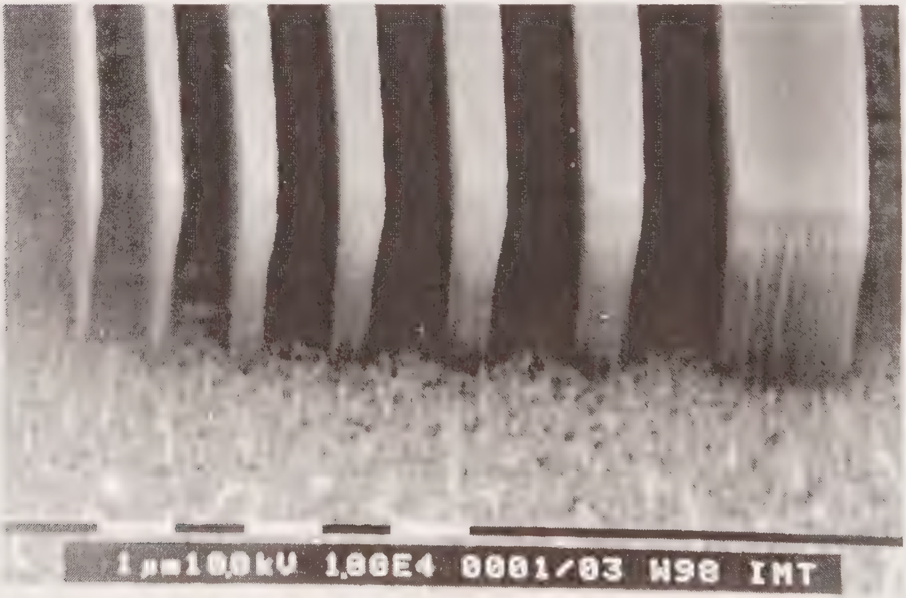

1 µm 10.0kV 1.00E4 0001/03 W98 IMT

FIG. 6.5 Resist structures made using reactive ion etching. Resist height is 3–3.5 µm and the smallest feature size is 0.3 µm.

After plating a 2–2.5-μm-thick gold absorber into the resist, the intermediate mask is used to transfer the absorber pattern into a 20–30-μm-thick PMMA resist using soft x-rays. Due to the high accuracy of x-ray lithography, hardly any structural losses occur in this step. Plating of gold in the developed sections results in the desired absorber thickness of 10–20 μm.

Alternative mask fabrication techniques, mechanical micromilling [16], and subtractive RIE [17] have been utilized to produce patterns in high-Z absorber materials. However, further development is needed before either of these approaches can become standard processes.

6.1.3.2 Membranes

A major part of an x-ray mask is the mask membrane. As already mentioned, there are different membrane materials possible. However, the requirements imposed on mask membranes are unique for all of them:

- high transparency to x-ray photons in order to reduce dose absorption in the membrane and maximize the dose deposition in the resist,
- high radiation resistance as high doses are accumulated during the long exposure times needed for patterning thick resist layers,
- high dimensional stability to ensure a high quality in pattern transfer,
- compatibility with thick resist layer application process and electroforming of gold, and
- optical transparency to allow aligned exposures.

In general, high transparency is a tradeoff between a low absorption coefficient for x-rays and membrane thickness while taking into account mechanical firmness and dimensional stability. Low Z materials such as beryllium, carbon (graphite, diamond), Si and Si compounds, and polymers offer the highest transparency. Of these materials, polymers and materials commonly used in semiconductor x-ray lithography offer poor radiation resistance and therefore limited lifetime and quality.

In Fig. 6.6, the x-ray transparency of various materials as a function of membrane thickness is shown. From these data, Be offers the best properties. Even several hundred micrometer-thick membranes from Be are superior to 2–5-μm-thin Si membranes. However, shifting the x-ray spectrum to shorter wavelengths is needed for patterning ultra-thick resist layers beyond 2 mm, which automatically requires both thicker filter absorbers to adapt the spectrum and thicker Si-based membranes [2, 18].

Since Be and Be oxides are highly toxic, the polishing of Be foil can only be done under special conditions, resulting in an increased cost for mask blanks. Alternative

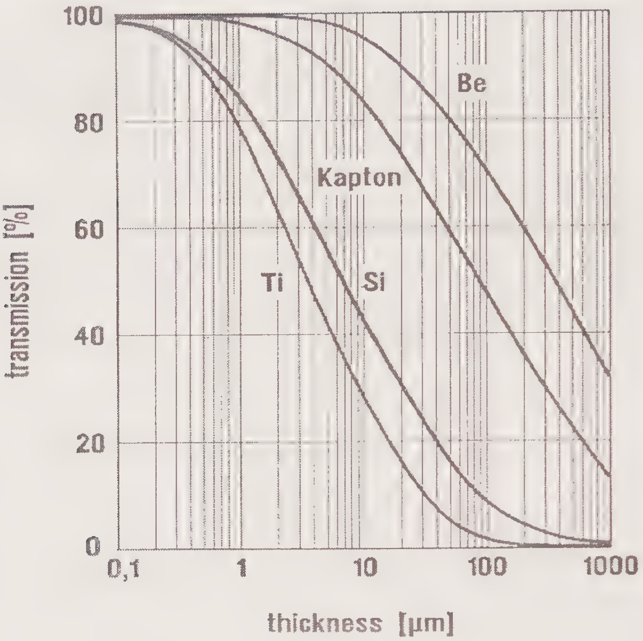


FIG. 6.6 Transparency to x-rays of typical mask membrane materials. Calculations are for Bonn University synchrotron at an electron energy of 2.0 GeV with a characteristic wavelength of 0.56 nm.

membrane materials such as 2–2.5-µm-thick Ti membranes [19], 4–5-µm-thick diamond membrames [6], 2-µm boron-diffused Si membranes [10], and relatively thick (125, 250-µm) graphite membranes [16] are in use.

The fabrication processes and transfer to a supporting ring are different for the thin membranes. Titanium membranes fabricated from sputtered layers are processed on a silicon wafer as a carrier and after plating of the absorber structure are lifted off to a frame made from Invar (alloy of 18% Ca, 28% Ni, 54% Fe), which is glued onto the membrane. Silicon membranes are fabricated by anisotropic etching using boron-diffusion layers as an etch stop and support rings made from Pyrex glass [10, 13]. Graphite membranes are used as sheet material and these are then clamped onto the resist surface [20]. Size and format of the membranes differ and haven't been standardized thus far.

The fabrication of microstructures with extreme heights of more than 1 mm is made possible by ultra-deep x-ray lithography (UDXRL) using "hard" synchrotron radiation with photon energies of 20 KeV and higher. When compared with a standard deep x-ray lithography (DXRL) process in which the radiation-sensitive resist film thickness is on the order of 1 mm or less, UDXRL has unique requirements that impose further restrictions on the choice of membrane and absorber materials. Recently, the first studies using hard x-rays of 20 keV photon energy to pattern centimeter-thick PMMA sheets have been carried out. This method is applicable to fabricating relatively large precision parts like ink jet nozzle plates [18]. The high photon energies needed to penetrate into thick PMMA sheets require different mask membrane properties and absorber thicknesses. 50-μm-thick Si wafers or 1-mm-thick Be sheets covered with Au absorber pattern as high as 50 μm are suitable masks for these conditions [21]. These "membranes" are readily available and the Au pattern can be defined by optical lithography, making this technology easy to use and relatively inexpensive. However, limitations in structural details as well as secondary effects enable only certain types of the smallest features to be patterned.

6.1.4 Processes for Deep-Etch X-Ray Lithography

In deep-etch x-ray lithography, the absorber pattern of the x-ray mask is transferred into a thick resist layer (from 100 μm up to one millimeter) by shadow printing using x-rays. Most of the impinging x-ray flux is completely absorbed in the thick absorber pattern. In the nonpatterned transparent mask regions, a large percentage of the x-rays that pass through the mask are absorbed in the resist layer, which is, in almost every case, polymethyl methacrylate. The result is a chemical modification of the resist in the exposed regions.

The accuracy limit of the pattern transfer is a function of the beam divergence, Fresnel diffraction from the absorber edges, and the penetration depth of photo electrons into the shadowed sections. Furthermore, secondary effects from fluorescence electrons produced in the mask membrane as well as photo and Auger electrons from the substrate will have an impact on accuracy. This is especially true for thicker resist layers and where there is a corresponding need for a harder x-ray spectrum.

6.1.4.1 Fabrication of thick resist layer

Because spin coating PMMA layers several hundred micrometers thick is not a suitable way to achieve thick homogeneous resist layers, other alternatives, such as casting and *in situ* polymerization [22], and glueing or solvent bonding of PMMA sheets [23] have been developed. The maximum thickness achieved by multiple spin-coating is on the order of 150 μm when using PMMA. Typically, a layer of up to 30 μm can be achieved in one step and with reasonable homogeneity across the entire substrate area.

The casting process starts with a highly viscous liquid formed from high molecular weight PMMA dissolved in methyl methacrylate. Following the addition of a hardener and an initiator, the liquid is poured onto the substrate and polymerized in a press. Spacer shims placed at the sides of the resist-covered area ensure that the appropriate resist thickness is achieved after polymerization. The molecular weight distribution of the polymerized resist strongly depends on the ratio of hardener and initiator used and can be either uni- or bimodular. A major problem related to this technique is the shrinkage in volume that occurs during polymerization. This shrinkage causes significient stress build-up, which leads to crazing and loss of adhesion. In order to minimize these defects, a very high molecular weight distribution is preferred, which is achieved by adding cross-linking and adhesion-promoting chemicals to the initial mix. In addition, a gentle annealing process applied after polymerization reduces the built-up stress. PMMA and the substrate typically differ in thermal expansion coefficient by an order of magnitude. Above the glass transition temperature (Tg) of the polymer the PMMA is a quasi liquid that easily adapts to any stress induced by a temperature change. However, below Tg, the PMMA becomes a solid and stress will build-up during a temperature change. To minimize the stress and give the PMMA material enough time to relax, an annealing program is used that ramps the temperature at 5° C/hr from Tg to approx. 50° C.

The glueing process uses the same material but not in the form of a thick resist layer. In this case, a thin (<10-µm) layer is applied to the substrate and a PMMA sheet is pressed into this layer. Polymerization takes place at room temperature using the same hardener and initiator system. Experiments done so far indicate that only a small amount of stress occurs in the thin glueing layer. Both methods are applicable to prestructured substrates because they easily adapt to the topology and deliver a planar resist surface. This is advantageous when combined processes are used [24].

The solvent bonding technique has been shown to be nearly free of built-in strain [18, 23]. In this process, a 1–2-µm-thick PMMA layer is first spin-coated on a planar substrate. After an annealing step, a few drops of solvent are put on this layer and a PMMA sheet is pressed onto the substrate. Drying of the solvent for several hours completes the process.

A major problem with high-aspect-ratio structures is maintaining good adhesion of the resist to the substrate following the x-ray lithography and development processes. Furthermore, the adhesion layer must provide a suitable plating base for the subsequent electroforming process. The cross section of this material for photo electrons also needs to be small in order to avoid dose deposition in the shadowed sections. These requirements can be fulfilled by using a sputtered titanium layer that is oxidized by a wet chemical process just prior to resist application. The porosity of the titanium oxide layer increases the mechanical bond to the resist. Adhesion can be further enhanced by utilizing internal adhesion promotors based on siloxanes, which chemically bond to the oxygen atoms of the metal surface.

For certain selected applications, other metal layers such as copper, nickel, and gold have been used as both adhesion and plating bases with reasonably good results. When a chrome/gold plating base is used for x-ray masks, chemically enhanced adhesion is made possible by exposing the layer to a thiophenol atmosphere prior to the application of the resist film by spin coating [6].

6.1.4.2 Radiation chemistry and resist development

One effect that has been observed in PMMA after x-ray exposure is a change in its chemical structure combined with a decrease in the molar mass [25]. Figure 6.7 illustrates the situation. Starting with a high molar mass (Mw= 650,000 g/mol) in the nonirradiated material, as the dose increases, this value is reduced to a minimum (Mw = 2,500–3,000 g/mol) for an accumulated dose of 20 kJ/cm^3. This reduction is caused by main-chain scission of the macromolecules. X-ray photons with energies in the kiloelectron volt range are absorbed by the individual atoms of the polymer due to the photo effect, and produce a range of high-energy photo and Auger electrons. These electrons continuously lose their energy, which leads to a great number of ionized and radiation-excited molecules as well as thermal electrons.

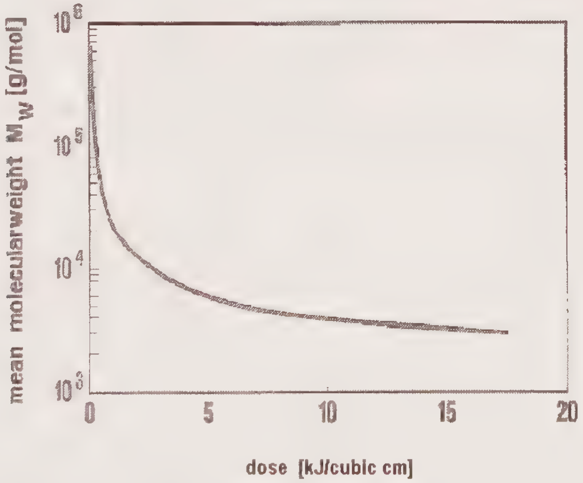


FIG. 6.7 Effect of x-ray energy dose on the mean molecular weight of PMMA.

Although the radiation chemistry of PMMA is a complex mixture of consecutive reactions among those radical species, main-chain scission is the dominant effect. The radiation-excited macromolecule splits off an ester side chain, producing an ester radical and a remaining chain radical. Stabilization of the chain radical is achieved by hydrogen abstraction and formation of a double bond, which results in a main-chain scission. Further recombination reactions of the remaining radicals result in stable molecules or trapped radicals.

A quantitative description of the radiation-induced reactions is given by the G-value, which denotes the number of main-chain scissions per 100 eV of absorbed radiation energy. At higher doses there are also cross-linking reactions in PMMA. This is due to the higher flexibility of the small chain molecules and therefore higher mobility, which leads to chemical bonds of chain radicals. In the state of equilibrium, scission and cross-linking reactions compensate each other, resulting in no change of the molar mass. This can be seen from Fig. 6.7 because there is no change of molar mass at high dose values.

After exposure, the irradiated regions have to be removed without any attack of the nonirradiated regions. A special mixture has been developed consisting of 20 vol% tetrahydro-1,4-oxazine; 5 vol% 2-aminoethanol-1; 60 vol% 2-(2-butoxy-ethoxy)ethanol; and 15 vol% deionized water [26]. The solubility of this developer with respect to the molar mass is shown in Fig. 6.8. Using this developer at 38° C allows the dissolution of the PMMA material to the depth of the microstructures (Fig. 6.8b) while there is hardly any attack of the nonirradiated regions, even at the top of the microstructures, which are exposed to the developer for some hours. The solubility of the developer can be shifted to a smaller molar mass by lowering the temperature. This increases the difference between the dissolution rates of the exposed and unexposed regions and allows for longer development times. A maximum temperature of approximately 45° C should not be exceeded because the high molar mass material in the unexposed sections starts to be attacked.

6.1.4.3 Exposure parameters for PMMA resists

PMMA is well known for its high resolution and accuracy properties as well as its low radiation sensitivity. This requires deposition of a relatively high dose and thereby long exposure time in order to reduce the molar mass to the required values defined by the developer. Furthermore, the material absorption in thick resist layers cannot be neglected. This is the reason for different absorption of doses along the resist height (see Fig. 6.8). Figure 6.8(a) shows a typical bimodal molecular weight distribution of a cast resist layer with a molar mass of 600,000 g/mol. The hatched area represents the molar mass, which can be dissolved by the developer at 38° C. As can be seen, the developer can only dissolve very little of the material, resulting in a high selectivity.

Figure 6.8(b) shows the unimodal molecular weight distribution achieved after exposing the PMMA to a dose of 4 kJ/cm^3. The molar mass of the distribution is 5,700 g/mol and it can be completely dissolved. A smaller dose will result in a higher molar mass, thus leaving some resist residues at the bottom of the microstructures. Therefore, this is a minimum value for the dose absorbed at the bottom of the resist layer. Nevertheless, smaller dose values can be accepted if higher temperatures and additional actions are taken, such as megasonic-enhanced development or stirring.

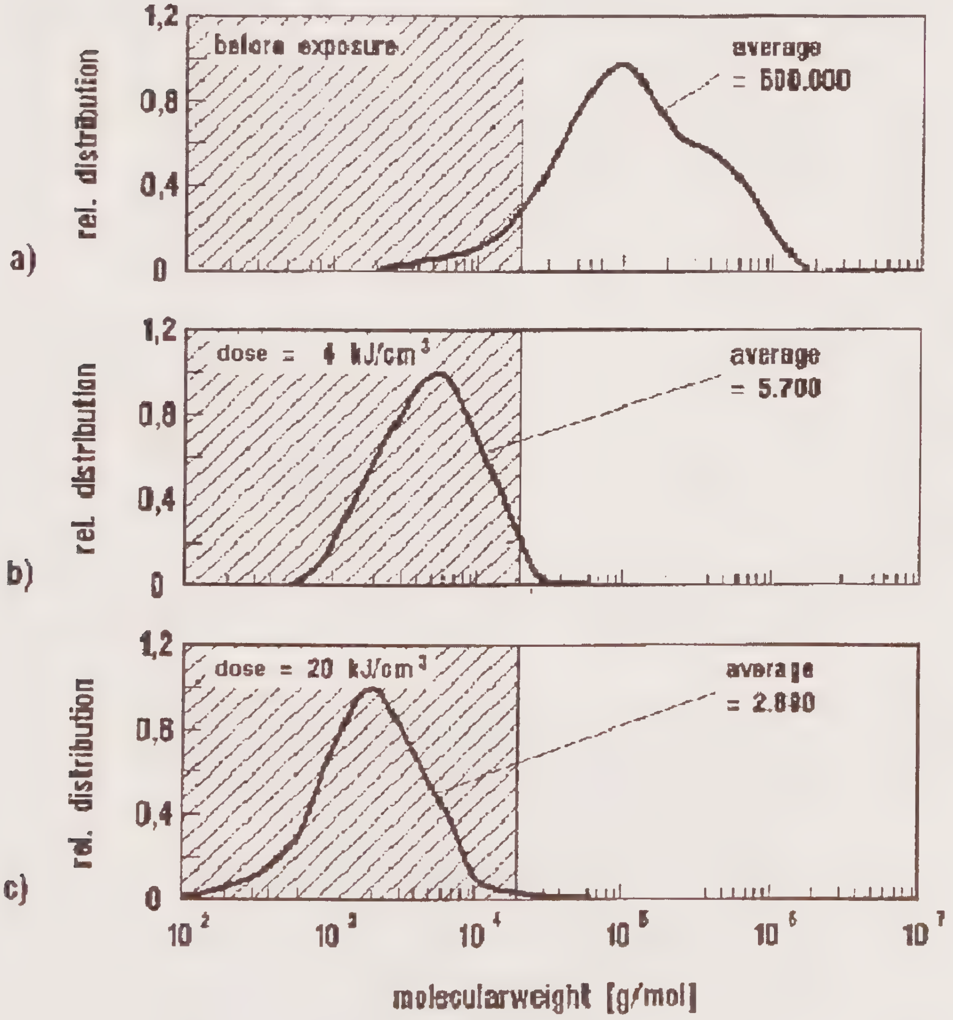


FIG. 6.8 Selectivity of developer on irradiated PMMA as a function of exposed molecular weight. The molecular weight distribution of PMMA before exposure, after depositing 4 kJ/cm^3 and 20 kJ/cm^3 of dose, is shown. The hatched area represents the molar mass from which 50% or more can be dissolved by the GG developer at 38° C.

Figure 6.8(c) shows the unimodal molecular weight distribution achieved from an absorbed dose of 20 kJ/cm^3, which is a typical value for the dose absorbed at the resist surface (top dose). The related molar mass is 2,800 g/mol. Higher dose values will result in resist foaming, which may lead to a loss in structural quality and therefore cannot be accepted. The chemical modifications going on with dose deposition not only lead to main chain scission but also involve the production of gaseous products. Due to the temperature rise, these products can be released from the PMMA matrix leading to bubbling or foaming. The height of the foamed resist can exceed several hundred μm and can be accepted if the proximity distance between resist surface and mask membrane is large enough [25]. However, this value has to be adapted to the temperature during exposure and to the resist thickness. For example, patterning of a 1-mm-thick PMMA structure reduces the top dose to approximately 14 kJ/cm^3 due to the low thermal conductivity of PMMA.

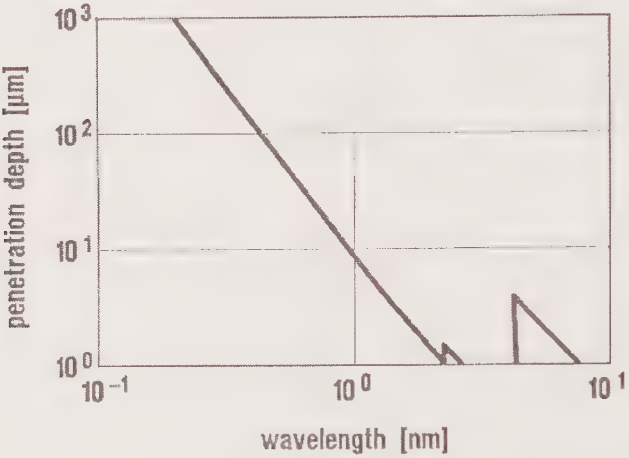


FIG. 6.9 Penetration depth of x-rays in PMMA as a function of wavelength (energy).

As a consequence, the exposure conditions for PMMA resists must take the limits for the top and bottom doses into account. The ratio should not exceed a factor of 5 to avoid foaming. This has an impact on the energy (wavelength) of the x-rays used for exposure. In Fig. 6.9 the penetration depth as a function of wavelength is shown in which the intensity at the bottom drops to one-fifth of that at the surface. In order to pattern a 500-μm-thick PMMA layer, a short monochromatic wavelength of 0.25 nm (photon energy of 5 keV) is needed. The lower energy (longer wavelength) x-rays are completely absorbed in the layers near the surface and do not reach deeper layers. While the synchrotron emits a broad spectrum instead of monochromatic light, additional filters may be inserted into the beam in order to adapt the spectrum to the required dose ratio.

Because high-energy x-rays can also penetrate through a high-absorbing material like gold, there is also some dose deposition underneath the absorber pattern. However, this critical dose value is typically below 100 J/cm^3. This results in a mask contrast of more than 299, which is related to a minimum gold thickness of 10 µm for a 500-µm-thick resist layer. Secondary effects may require an even thicker gold absorber.

6.1.4.4 Accuracy limits

The main qualities of LIGA microstructures are high aspect ratio and nearly vertical side walls along the entire height of the structure. As an example, a cross-bar microstructure is shown in Fig. 6.10(a). The width of the bars as a function of height has been measured and the result is shown in Fig. 6.10(b). Linear regression analysis indicates nearly perfect vertical side walls, with a slope of only 0.04 µm per 100 µm of resist height.

The contributions to any measured deviation from a perfectly vertical side wall are manyfold. For example, inaccuracies in scanner motion during x-ray exposure, high (but not perfect) selectivity of the developer, or a sloped absorber side wall profile are specific effects related to the quality of the process and as such are more readily controlled. The more fundamental limits, however, are related to the primary effects resulting from the x-ray exposure. Natural divergence of the synchrotron source, Fresnel diffraction at the absorber structures, and penetration depth of photo electrons [7], as well as secondary effects from fluorescence produced by the mask membrane and the substrate material and mask distortion due to heat problems in the mask [7], define a more fundamental accuracy limit of the process.

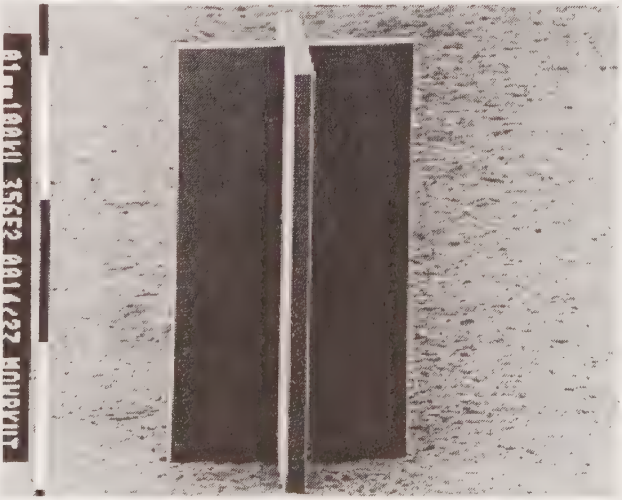


FIG. 6.10(a) Example of a vertical straightness attainable by deep x-ray lithography.

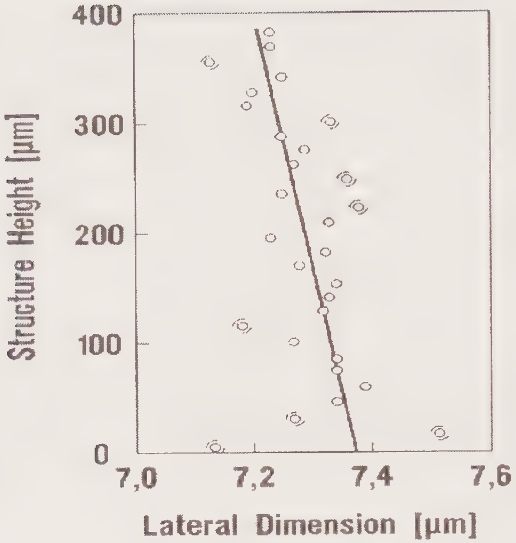


FIG. 6.10(b) Lateral deviation of test structure as a function of height.

6.1.4.5 Fresnel diffraction and photo electrons

The absorber structures on the mask diffract the incident x-ray photons, resulting in a smeared dose deposition along the edge of the structure, namely, in the absorber-covered regions. This dose variation is a function of the photon wavelength, the proximity gap, and the lateral position from the absorber edge. In the case of monochromatic radiation of wavelength λ, energy is deposited at a maximum distance $x_{\max}$ relative to the patterned edge and to a depth z from the resist surface according to the mathematical relationship [27],

$$x_{\max} \approx \sqrt{\lambda \frac{z}{2}} \quad .$$

As can be seen, this effect decreases with shorter wavelength or higher photon energies. It is even smaller for polychromatic radiation from a synchrotron source because no completely destructive interference occurs. Taking into account an ideal developer, the effect of Fresnel diffraction on the side-wall accuracy has been calculated for a 500-µm- thick resist and is shown in Fig. 6.11 as a function of the characteristic wavelength. The deviation from the ideal side wall is calculated from the maximum distance $x_{\max}$ where the first intensity minimum appears. While DXRL needs an x-ray spectrum with photon wavelength below 1 nm, the lateral deviation caused by diffraction is on the order of 0.1 µm.

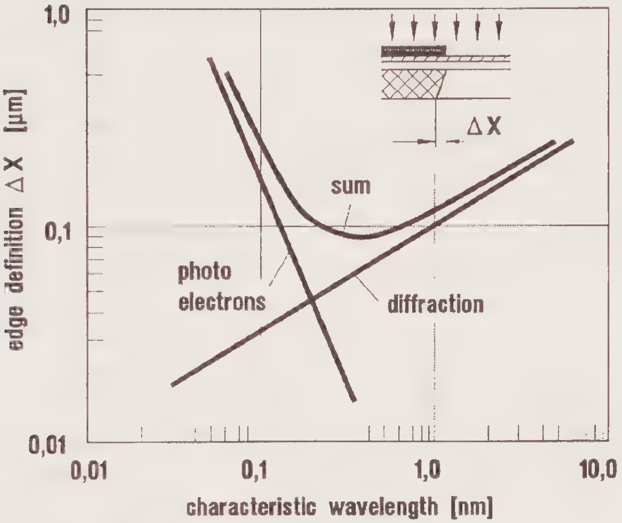


FIG. 6.11 Effect of edge diffraction and photoelectrons on the edge acuity of a 500-µm-high resist structure assuming a perfect developer [6, 7, 119].

In addition, the absorbed x-ray photons create photo and Auger electrons that interact with the resist material, resulting in a chemical modification of the resist. Due to the high photon energy and the isotropic distribution, these electrons are able to penetrate into the shadowed resist regions. The corresponding dose deposition also reduces the dimensional accuracy of the side wall. A rough estimate of the edge definition due to the photo electrons is given by the relation shown in Fig. 6.11. The deviation is proportional to $1/\lambda^2$.

The impact on the lateral deviation is also shown in Fig. 6.11. In contrast to Fresnel diffraction, the penetration depth increases with shorter wavelength. The total lateral deviation is the sum of both effects and shows a minimum of 0.1 µm for characteristic wavelengths between 0.2 and 0.3 nm.

6.1.4.6 Divergence of synchrotron radiation

Synchrotron radiation is emitted in a very small angular cone relative to the plane of the electron trajectory [28]. The opening angle or natural divergence is mainly given by the electron energy. Using synchrotron sources with electron energies of 1–2.5 GeV, the divergence is on the order of 0.2–0.3 mrad.

Beam divergence is also present due to the betatron oscillation of the electron ensemble, indicating that the electrons oscillate around the “ideal” or “reference” orbit [29]. The divergence from betatron oscillation is a machine-specific value and is further dependent on the location of the “source point” inside the ring lattice.

Typical numbers are 0.1–1 mrad. Furthermore, the finite electron beam size contributes to the overall divergence.

In order to qualify the impact of beam divergence on the accuracy of the side wall, the distance from the scanner to the source (typically 10–30 m) has to be taken into account as well as the total filter thickness needed to adapt the spectrum to the required exposure conditions. Typical values are 0.5 mrad for the vertical direction (direction of the scanner motion). In the horizontal direction, this effect is approximately one order of magnitude smaller and can be neglected [6].

6.1.4.7 Tilt of absorber side walls

A tilt of the absorber side walls relative to the incident beam is caused by two effects: (1) a misalignment of the mask normal with respect to the incident photons and (2) an inclined side wall of the absorber pattern itself. While the first effect can be controlled experimentally to better than 0.1 mrad by means of an optimized scanner alignment [6], the latter is a function of the mask fabrication process. This is especially true for masks patterned by optical lithography where side wall angles of 85 degrees are possible. The result is the generation of a finite transition region between exposed and unexposed regions [30].

The calculated lines of constant dose deposition underneath such a sloped absorber side wall using typical exposure conditions for DXRL experiments result in a slope in the resist pattern being approximately 200 times smaller. Assuming a maximum side wall angle of 85 degrees, the corresponding angle of the microstructure is approximately 0.8 mrad. The conclusion is that extremely vertical side walls cannot be achieved with x-ray masks made by using optical lithography. Therefore, intermediate x-ray masks must be used.

6.1.4.8 Mask distortion

Thermoelastic deformations of the mask membrane during exposure can result in a shift of the absorber side wall relative to the original position. Parameters affecting this distortion are scan velocity, area of the membrane covered by absorber, thermal and mechanical properties of the mask material, power density of the synchrotron source, and cooling of the mask during exposure. Experimental and theoretical results achieved so far indicate that high scan velocity and the use of thick Be foil as a mask membrane can limit the deformation to below 0.2 μm [31].

6.1.4.9 Fluorescence radiation from the mask membrane

The absorption properties of any material are dominated by the specific absorption edges. At these edges, electrons can be extracted from the inner shell of the atom and the absorption increases in a stepwise fashion by one order of magnitude. Recombination of this activated level produces fluorescent radiation, which is emitted isotropically [39]. The microstructure quality at the top of the resist is affected because an additional dose is deposited underneath the absorber pattern. Depending on the exposure conditions and on the resist thickness, a sufficient dose can be absorbed so that the developer is able to dissolve the resist. The resulting defect, a rounding of the side wall at the top, is shown in Fig. 6.12. In this case, a Ti mask membrane was used to pattern a 1-mm-high microstructure. Since the k-edge of Ti (5.4 nm) corresponds well with the maximum of the emitted synchrotron spectrum, a large radius of curvature of approximately 150 μm is achieved. Replacing the Ti mask with a Be mask and exposing the same structure leads to a very small radius of curvature of only 2 μm. This effect, combined with the inherent rigidity and high thermal conductivity that minimizes distortions caused by heat loading, shows beryllium would be the material of choice for the x-ray mask membrane in DXRL. However, beryllium exhibits oxidation problems that need to be resolved. An alignment window strategy also has to be worked out.

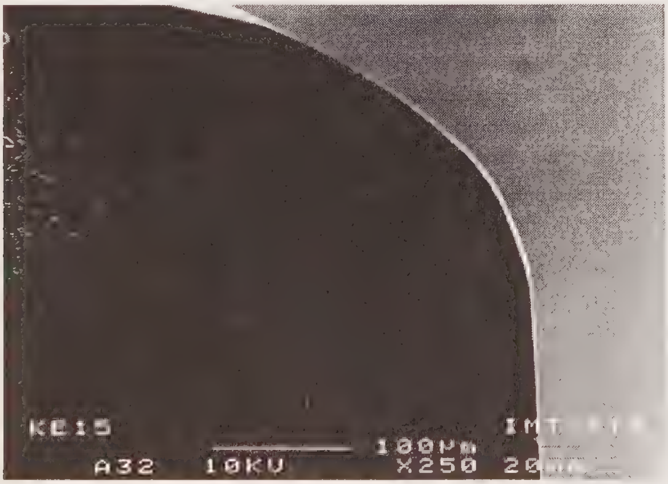


FIG. 6.12 Rounded edge of top of exposed structure due to fluorescence radiation from x-ray interaction with mask membrane. Exposure made at ELSA storage ring at Bonn University at 2.3 GeV (0.5 nm wavelength) with a 700-μm-thick polyimide filter and a 2.2-μm-thick titanium mask membrane. Resist thickness was 1.1 mm.

6.1.4.10 Secondary electrons produced in the substrate layer

In order to allow electroplating, the substrate surface is typically covered with a thin metal layer of Ti, Au, Cu, or a combination of these materials. High-energy photons

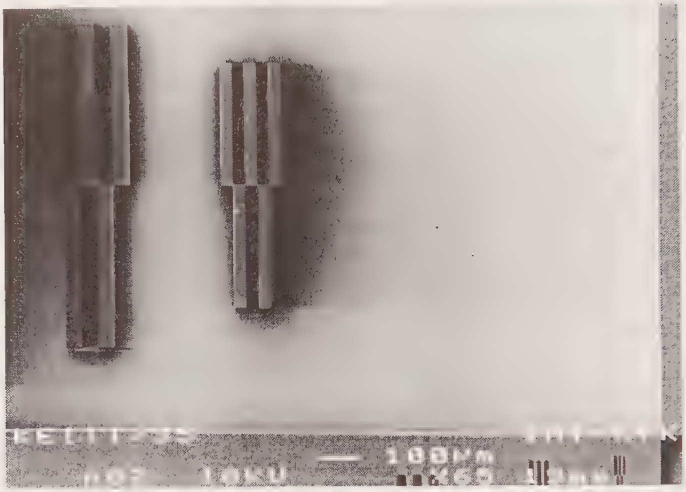


FIG. 6.13(a) Exposure with hard x-rays (critical wavelength of 0.2 nm) leads to adhesion failure of small structures.

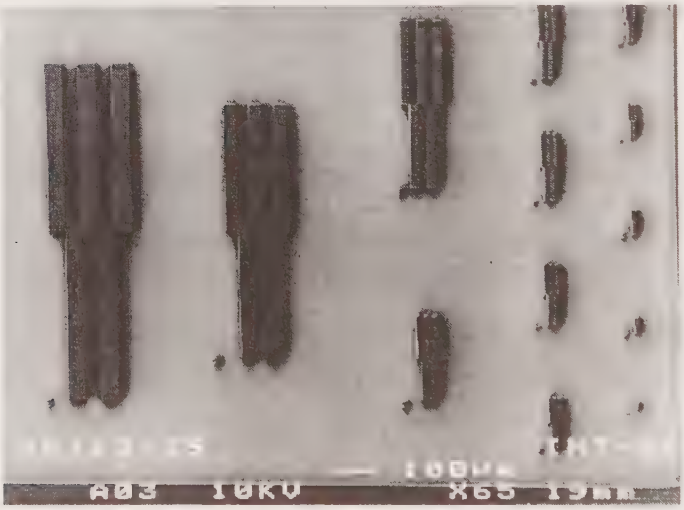


FIG. 6.13(b) Exposed with soft x-rays (critical wavelength of 1 nm) leaves good adhesion of small structures.

penetrating through the full resist thickness are efficiently absorbed in this layer. Secondary electrons (photo and Auger electrons) are produced with a high yield, and being emitted isotropically, create an additional dose deposition in both the exposed and unexposed regions. In case of the nominally unexposed areas, this may have a dramatic impact on the adhesion of the microstructure to the substrate, as shown in Fig. 6.13.

The scanning electron microscope (SEM) pictures show the same 100-μm high structures exposed with different conditions, namely, synchrotron spectra of different characteristic wavelengths. With a relatively soft spectrum (λc=1 nm), even with the smallest structures there is no loss of adhesion to the substrate. On the other hand, a hard spectrum (λc=0.2 nm) leads to a complete failure even for the larger structures. This indicates that the hard x-rays, which are able to penetrate through the Au absorber, increase the dose deposited at the resist–substrate interface. As a consequence, for hard x-ray spectra, the minimum absorber thickness has to be much greater in order to compensate for the additional dose contributed by the secondary electrons. Another approach may use x-ray mirrors under grazing incidence in order to absorb the very hard part of the spectrum [9, 32].

Although the previously described effects have been shown to influence the quality of the pattern transfer process using DXRL, the overall quality of the structures fabricated with DXRL is truly extraordinary, especially when compared with structures produced by other fabrication methods. Further process enhancements can minimize these effects to the point where DXRL will provide a significant advantage over other techniques for fabricating microstructure components for mechanical, optical, and fluidic applications.

6.1.5 Modified LIGA Processes

Applications in microsystems not only require accurate microstructures but also advanced assembly of components, interconnection within the system, and interconnection to the macroworld. In order to fulfill these requirements, the LIGA process has been modified and further adapted to the system applications. The following section shows some examples of modified LIGA processes and illustrates the possibilities for system integration.

6.1.5.1 Sacrificial layer technique with the LIGA process

The fabrication of micromechanical sensors and actuators requires both movable components and microstructures to be fixed to the substrate. Since the components are typically interconnected with each other and the tolerances are very small, assembly procedures range from extremely difficult to virtually impossible. However, by combining the LIGA process with a sacrificial layer technique, this limitation can be surpassed. As an example of this modification, the fabrication steps of a LIGA acceleration sensor are shown schematically in Fig. 6.14.

In this case the movable and fixed parts have to be electrically insulated. Therefore, the process begins with an insulating substrate, such as a silicon wafer with an oxide layer or a ceramic substrate. Using PVD processes, a metallization layer is deposited onto the substrate. The metallization typically consists of an adhesion layer to guarantee good adhesion to the substrate and a conductive layer to serve as a plating base for the electroforming process. This multilayer system is patterned by a combination of optical lithography and wet etching processes.

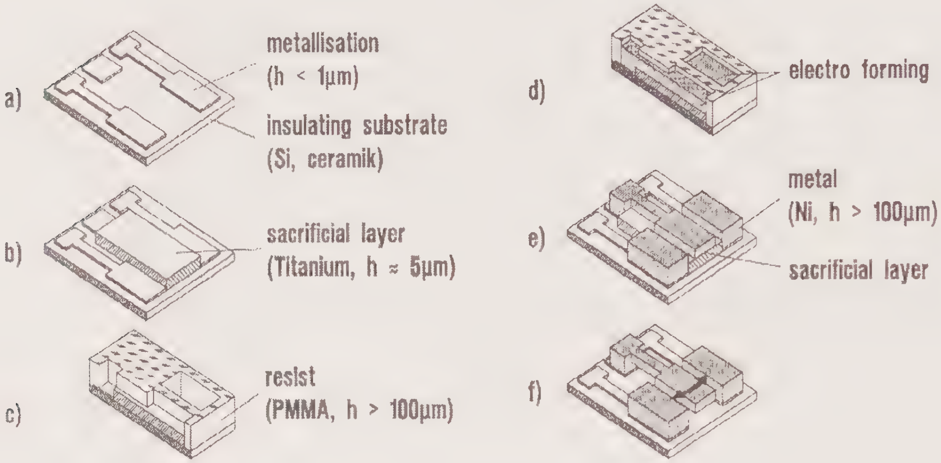


FIG. 6.14 Process schematic of LIGA utilizing a sacrificial layer.

Next, the sacrificial layer is deposited onto the previously prepared substrate by a PVD process. For the LIGA process, a suitable sacrificial layer is titanium. This material provides excellent adhesion to both the resist and the conductive layer as well as etch selectivity to nickel and other metals used with electroplating. The titanium is then patterned by optical lithography and etching. In the aligned x-ray exposure, the parts that will be moving are patterned onto the titanium-covered surface while the fixed parts are patterned directly on the conductive layer.

Other sacrificial layers such as Al are also used in the LIGA process. The advantages of Al over Ti are lower internal stress and no shadows upon removal [33]. Nevertheless, more uniform plating is achieved by covering the Al with a 20-nm-thick, thermally evaporated capping layer of Ti.

The aligned x-ray exposure requires special alignment marks in the mask membrane [19]. The mask alignment marks are adjusted to marks patterned on the substrate with an optical microscope to achieve an overlay accuracy of 1 µm or better.

After electroplating and stripping of the remaining resist, the Ti layer is selectively etched with 0.5% HF. The movable parts are released from the substrate while the fixed parts are anchored to the conductive layer. The individual microstructures are electrically connected to an electronic circuit board by wire bonding on special bond pads, which are part of the conductive layer system.

CMOS electronic circuits that are fabricated on an Si substrate cannot be used with x-ray lithography due to radiation damage to the microelectronic circuits. However, a direct integration of LIGA microstructures with a CMOS wafer is possible using aligned molding techniques [34].

6.1.5.2 3D Patterning in deep x-ray lithography

Typically, LIGA structures have an arbitrary 2D pattern that is extended into the third dimension to form prismatic or cylindrical geometries. Geometrical variations in the third dimension are possible and can be obtained in different ways. Stepped microstructures can be achieved either by micromachining a substrate and adding a LIGA plane on top [35], or by multiple, aligned LIGA processing [18, 36]. Tilting and/or rotating of the mask/sample assembly during exposure [7, 37] leads to LIGA structures with sloped side walls. An additional annealing step after development [38] can produce spherical surfaces on top of LIGA structures while making use of fluorescent radiation from the mask. Additional filter membranes placed between the mask and resist surface allow patterning of conical structures [39].

Multilevel stepped LIGA microstructures. There are different methods to create a multilevel microstructure. In the simplest approach, a coarse microstructure is first exposed to a resist layer. This is followed by a second aligned exposure containing fine structures. In the second exposure, the minimum bottom dose required to completely dissolve the exposed resist is deposited at some predetermined depth and not at the surface of the substrate. During development, the coarse structures are developed down to the substrate surface while the fine structures are only developed to a certain depth, thus forming a level at a different height.

Another approach is based on the consecutive formation of individual resist layers and the separate exposure of each layer [36]. The first layer is either cast or solvent bonded onto a substrate and then exposed to form the first latent image. A second resist layer is solvent bonded over the first layer and is exposed with an appropriate mask which forms a latent image in this layer. This process can be repeated several times and multilevel structures can be achieved. The accuracy for each step is determined by the accuracy of the individual resist sheets. After finishing the exposures, the 3D image is formed by developing the sample in one step. Nevertheless, since x-rays don't have a defined stopping depth the underlying resist also gets partially exposed, which results in fairly rough surfaces.

In principle, using multilevel or even analog x-ray masks with varying absorber thickness can lead to similar results and offers the advantage of only one exposure step and no alignment in order to generate the complete pattern. However, the problem of fabricating such masks hasn't yet been solved.

An alternative approach is based on a combined process of exposure, electroforming, and polishing [18]. In this case, a first level is fabricated by plating a metal into the developed pattern geometries and polishing the metal/polymer composite to the desired height. A special lapping/polishing procedure has been developed that delivers a well-controlled planarization capability with a tolerance of ±2 μm and low surface roughness. A second layer is solvent bonded to the composite and patterned with a second aligned x-ray mask. A repeat of the LIGA cycle generates the second metal level. However, the accurate lateral alignment of the various levels is an issue that needs to be addressed in the future.

The last method being investigated is the combination of processes such as LIGA and Si micromachining [24]. In this case an Si wafer is used as a substrate and a coarse pattern is generated, for example, by silicon wet etching. Fine tuning of the height of this first level to better than 1 μm is possible by additional sputter etching. The PMMA resist is cast onto the stepped wafer, which is then used as a substrate in x-ray lithography. The mask pattern is transferred to the substrate surface at each point and the pattern is developed down to the surface. In this case the step height is defined by the accuracy of the first process and can be controlled to approximately 0.1–0.2 μm. Due to the design flexibility of the individual layers, overlay tolerances of the LIGA mask and the prestructured substrate are only a few micrometers, which can easily be achieved.

Instead of using prestructured Si substrates, it is possible to use machined metal substrates to fabricate multilevel structures. Furthermore, the PMMA resist can be structured by hot embossing using a mold insert produced by LIGA or another process to generate a first level [40]. In both cases, an aligned x-ray lithography process will generate the second level.

In conclusion, various methods of creating multilevel structures are available. The requirements and dimensional tolerances are application specific and will determine which method is the most efficient.

Microstructures with sloped side walls. In order to pattern microstructures with a side wall orientation other than 90 degrees with respect to the substrate surface, the tilting of the mask/substrate relative to the x-ray source has been successfully used. Also, there is a transition region on the x-ray mask where the absorber thickness continuously increases from zero (no absorber) to its maximum. A sharp, well-defined side wall is achieved in the resist due to the high selectivity of the resist/-developer system [7]. Combining multiple exposures at different angles with a predetermined shift between mask and resist during consecutive exposures also allows more complex 3D-like microstructures to be fabricated by x-ray lithography [41].

Conical and spherical LIGA microstructures. Taking advantage of secondary radiation effects offers an alternative method of 3D patterning. Placing a radiation filter between the mask and resist surface generates secondary electrons and fluorescent photons, which are emitted isotropically. By controlling the type of filter material, the synchrotron radiation spectrum, the distance between the membrane and the surface, and the exposure dose, a limited patterning of the shadowed sections near the surface is possible, leading to a conical shape [39].

Spherical shapes of microstructures can be achieved by another process modification [38]. After exposing and developing cylindrical microstructures, these structures are exposed a second time without a mask and with a relatively soft spectrum This leads to a dose deposition mainly in the upper part of the microstructure. Along with the absorbed dose, the molar mass as well as the glass transition temperature (T_g) is changed. As a consequence, the glass transition temperature varies with height and is lower at the top of the microstructure. An annealing step at a temperature slightly above the T_g of the top material will transfer the material in a quasi-liquid phase while it is still in a solid phase at the bottom part. Surface tension then forms the top of the microstructure into a spherical shape while the solid bottom part ensures the dimensional stability. This process can be used to fabricate high-aperture microlenses and microlens arrays.

Light-guiding microstructures fabricated by hot embossing. For applications in fiber optic networks or in fiber-based sensors, polymer waveguide structures are attracting more and more interest because they offer the possibility of mass fabrication. However, dimensions and requirements on tolerances differ greatly between single and multimode waveguide structures.

The LIGA process allows the parallel fabrication of the waveguide system, fiber fixing grooves, and other optical components, such as reflection gratings, in one step, thereby reducing alignment efforts to a minimum. As an example, a microspectrometer that consists of a multimode waveguide system with a combined reflection grating has been fabricated. This was accomplished by combining hot embossing of polymer sheets with slightly different refractive indices with polymer welding [42] (for details, see Section 6.3).

6.2 MICROMECHANICAL MACHINING PROCESSES

6.2.1 Introduction

The micromechanical machining processes are those that utilize direct removal of the workpiece material by means of a physical cutting tool such as a microdrill or a diamond turning tool, or an energy-related process such as laser ablation or electrical discharge machining. The workpiece may be a final part, a mold for electroplating to create a microstructure, or an x-ray lithography mask, for example. The difference between lithographic and micromechanical processes is the manner in which material is removed to form a feature rather than in the function of the finished structure [43].

Micromechanical processes are often thought of as conventional machining processes which have been merely shrunk down to work at the micrometer scale. While this is partially correct, micromechanical processes require specialized tools and methods that cross over into the realm of precision engineering. These methods also include vibration isolation and control, thermal stability and lack of thermal gradients in parts and tools, and the approach that errors are not random but happen because of conditions that can be corrected or at least partially controlled or compensated [44].

However, micromechanical processes often violate one underlying definition of precision engineering: that the part or feature tolerance is 1 part in 10,000 of the part or feature size. As an example of this, a micro endmilled structure is shown in Fig. 6.15. The trench is nominally 23 μm wide with a standard deviation of 0.3 μm. By the definition of precision engineering, the tolerance (effectively ± 3 standard deviations) on the trench width would need to be 2.3 nm. Such a tolerance is impractical or impossible to fabricate and for most foreseeable structures and applications, such a tolerance would not be needed [45].

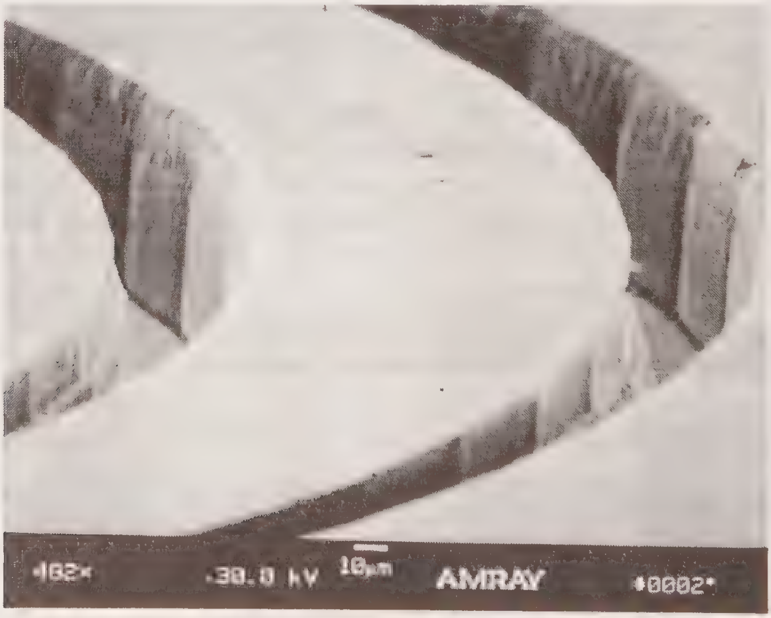


FIG. 6.15 Portion of micromilled structures in PMMA using 2-fluted endmill at 19,000 rpm.

A second deviation from merely miniaturizing conventional processes is in the cutting tool design. A miniaturized four-fluted end milling tool is shown in Fig. 6.16. Although the figure shows the degree to which miniaturization is possible, such a tool cannot perform the milling process due to adhesion of the work material in the small crevices at the base of the cutting edges. In addition, although the cutting edges are as sharp as many conventional cutting tools, the edges are very dull compared with the size of the workpiece chip that is machined with the tool. This represents one of the primary differences between micromechanical machining and miniaturized conventional machining. In conventional machining, most material is removed by shearing along a plane partially determined by the primary rake face of the cutting tool [46,47]. Cutting edge sharpness plays a less important role than in micromechanical machining in which all of the material shearing takes place along a relatively dull cutting edge (the uncut chip thickness is approximately equal to the cutting edge radius), which decreases the effective rake angle, resulting in high cutting forces relative to the depth of cut [48]. This, coupled with the fact that the cutting tools are inherently weak due to the small size involved, makes machining at this scale a significant challenge.

Because the micromechanical processes directly remove material in one step, there can be less buildup of dimensional errors than with serial-type processes. The micromechanical processes are sometimes termed complementary processes

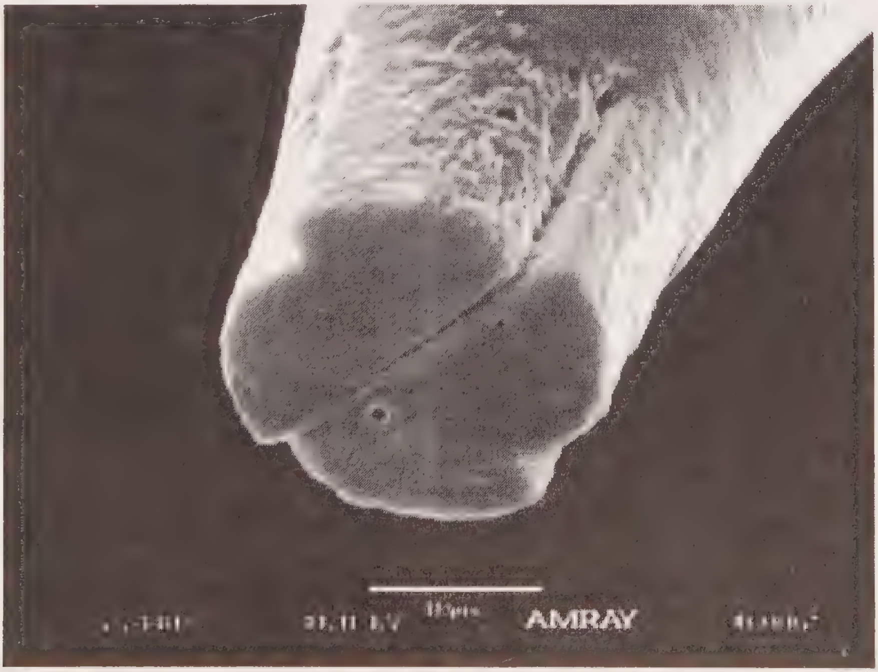


FIG. 6.16 Four-fluted micromilling tool made of high-speed steel. Cutting edges are rounded due to Gaussian energy distribution of ion beam used to fabricate the tool.

because they can be integrated with other processes (lithography or plating, for example) to enhance the effectiveness of all the micromanufacturing processes. The micromechanical processes are not competing processes but do have a place during prototype development and for applications where the micro-to-milli size ranges are integrated. The following sections provide a concise overview of the most applicable micromechanical processes and how they can be used to fabricate or enhance the fabrication of high aspect ratio structures.

6.2.2 Diamond Machining

Many years of research and development have been devoted to improving the quality of cutting tools. Most of this work has been in the area of materials development to create tools which are harder, sharper, have better thermal properties, and will maintain size and shape under the influence of cutting forces. Single-crystal diamond is such a material. A comparison between diamond and high-speed steel is shown in Table 6.1.

TABLE 6.1 Comparison of diamond with high-speed steel.

Property	Diamond	High-Speed Steel
Shear modulus (GPa)	300	79
Elastic modulus (GPa)	1000	206
Critical tensile strength (GPa)	4	4
Room temperature thermal conductivity (W/(m * K))	1600	60
Maximum cutting temperature (°C)	1600	650

The high elastic modulus of diamond results in decreased straining and deformation under the influence of cutting forces. Diamond retains its high strength even at high temperatures, which can be experienced with the high cutting speed of a thermally insulating workpiece material, such as plastics. Diamond also has the highest thermal conductivity of any material at room temperature. Although this conductivity decreases at elevated temperature, much or nearly all of the heat generated in a cutting process (typically 80% or more for metals and 90% or more for plastics) will be carried away by the diamond. This means that the workpiece will experience less thermal growth, resulting in better precision of the process.

Perhaps the greatest advantage of diamond is its crystalline structure, which results in very sharp cutting edges. The theoretical sharpness of a diamond cutting edge has been estimated to be on the order of several nanometers. Although such an edge is possible, its use in an actual cutting operation could be questioned. Diamond tools with an edge radius of 20 nm have been used for machining purposes.

Many materials which are not considered diamond machinable at the macroscale can be effectively machined at the microscale. The primary reason for this is the relative amount of material removed. At the macroscale, a diamond tool will wear quickly relative to the volume of work material removed and therefore the process is not considered cost effective. The wear is mostly diffusion of the diamond into the material if the work material readily absorbs carbon, as does iron [49]. At the microscale, however, the volume of material removed is many orders of magnitude smaller and therefore a comparably sized (or smaller) diamond tool will be able to produce many more parts before the tool has worn. This can make diamond machining of these materials more cost effective at the microscale. Examples of materials regarded as both machinable and nonmachinable are shown in Table 6.2.

TABLE 6.2 Material compatibility for diamond machining.

Machinable Metals	Machinable Plastics	Machinable Crystals	Difficult to Machine Metals
Copper and alloys	Propylene	Germanium	Nickel and alloys
Aluminum and alloys	Polymethyl methacrylate	Silicon	Beryllium and alloys
Gold	Styrene	Lithium niobate	Ferrous alloys
Silver	Polycarbonate	Zinc sulfide	Stainless steel
Tin	Fluoroplastics	Gallium arsenide	Titanium and alloys
Electroless nickel		Cadmium telluride	Molybdenum and alloys

The materials that are considered nonmachinable may be machined at the microscale with some success. However, several of these materials have been successfully machined at the macroscale by cooling the workpiece and diamond tool to cryogenic temperatures [50]. This greatly reduces the rate of diffusion between the workpiece and the diamond.

At the microscale, however, the volume of material that is normally removed per part is very small and the tool can generally survive the production of more parts, thus making the process more feasible. If crystals are to be machined, the depth of cut into the work must be generally maintained in the nanometer range through the tendency to produce subsurface cracking because of crack propagation caused by tensile stress in the material. These special cutting conditions are termed "ductile-regime machining" and are specific to the type of crystal being machined [51,52]. The technique is to limit the maximum tensile stress in the crystal to a level below the critical stress intensity for a crack to propagate in the work material.

General guidelines for diamond machining conventional materials are shown in Table 6.3. These are only guidelines and should be refined for a specific machining application.

TABLE 6.3 General macroscale diamond machining guidelines.

Material	Surface speed (m/min)	Rough depth of cut (microns)	Finish depth of cut (microns)
Soft metals Al, Cu	100 - 130	50	2
Hard metals Ni	100 - 130	10	1
Plastics	100 - 130	250	12
Crystals	100 - 130	125	2*

* Highly dependent on stresses generated in material and critical stress intensity factor of material

Guidelines for diamond machining, such as those given, are adequate for "larger" microstructures such as those shown in Fig. 6.17. Each of these structures was machined with a conventional air-bearing-equipped lathe. The spindle rotated at 800 rpm and the depth of cut, etc. were according to the recommendations. These copper structures are microscale heat transfer surfaces appropriate for use in very high-flux thermal devices such as microcompact heat exchangers, heat pipes, chemical reactors, and multiphase thermal applications. Using surfaces such as these, heat fluxes as high as 300 MW/m^3K (log-mean temperature difference) have been obtained [53]. At the microscale, however, these machining conditions may need to be drastically altered, depending on the specific machining application. A diamond-machined microshaft is shown in Fig. 6.18. For this application, the spindle speed was 4,800 rpm and the typical depth of cut was as small as 0.5 μm for the finishing cuts. These changes in the machining parameters are required because of the slow surface speed of such a small component, and the great flexibility and low strength of a structure with a diameter of 25 μm and a length of 100 μm. To machine a structure as small as that shown, the radial depth of cut must be adjusted due to the fact that as the shaft is being machined, it tends to deflect away from the cutting edge of the diamond tool. In addition, the shaft attempts to rotate up and onto the rake face of the cutting tool, which will cause the shaft to break. This problem is

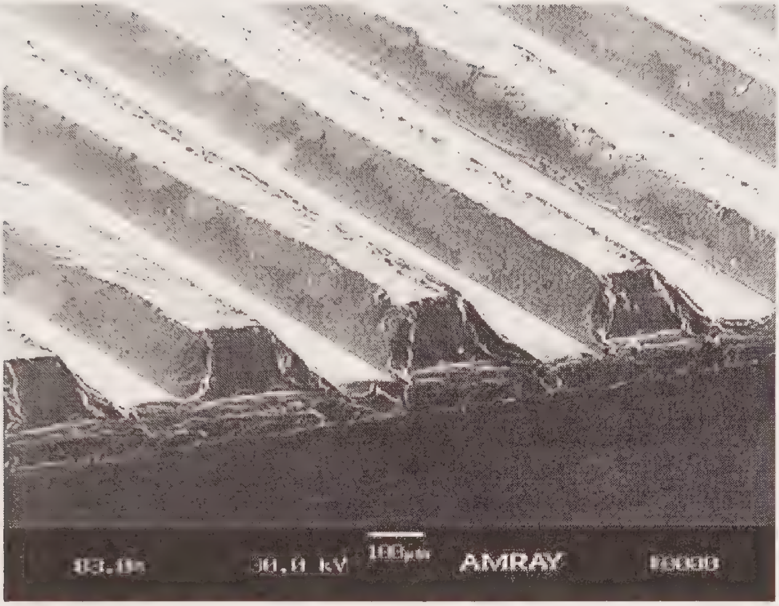


FIG. 6.17 Portion of diamond-machined microcompact copper heat exchanger plate. Layers are then vacuum-diffusion bonded to form complete heat exchanger.

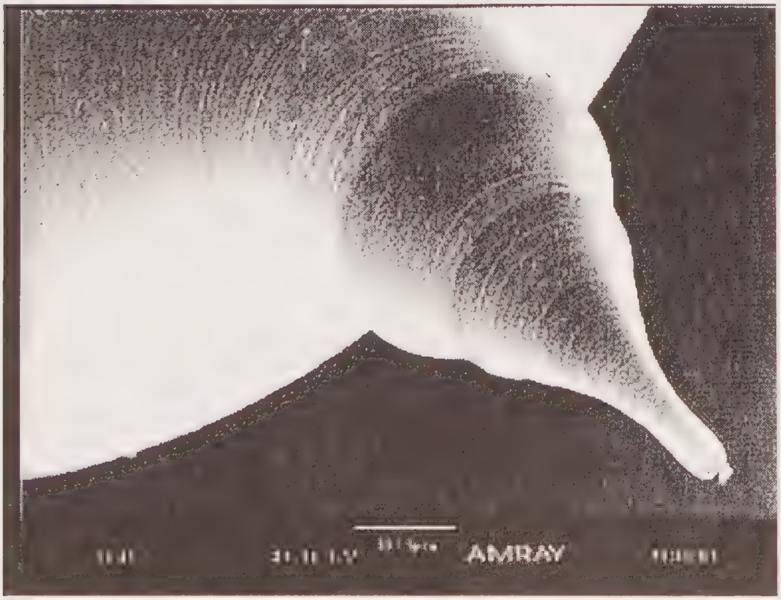


FIG. 6.18 Twenty-five-micrometer diameter by 100-µm-long microshaft made by diamond turning of A1-4303 with a polycrystalline diamond tool.

familiar to any machinist who makes parts with a very large length-to-diameter ratio (L/D) on a lathe.

Diamond machining, particularly at the microscale, requires extremely smooth movements in both the spindle and tool positioners. For this reason, fluid-bearing spindles and stages, or very high precision mechanical bearing spindles and stages are required. The stages must have excellent straightness and flatness to maintain a sufficiently accurate relationship between the tool cutting edge and the rotating workpiece. In addition, isothermal conditions are needed to maintain the precision of all the machining components. At the microscale, and at high precision, the most detrimental factor is often thermal instability or the lack of thermal control. High-precision machine tools are normally quite massive to help reduce transmitted vibration. Unfortunately, environmental control systems normally cycle at low frequency (many minutes). This low-frequency input allows a machine tool with a large thermal mass to follow the input reasonably well. If the environmental control system cycles, it should be set to cycle at a relatively high frequency so the massive machine tool does not have time to respond and deform. Determining the period of this fluctuation requires an analysis for a specific set of conditions.

Diamond machining can be used to create very smooth surfaces in materials because of the sharpness and stability of the cutting edge. This has particular application if the work material is PMMA, which is being used as a thick photoresist for deep x-ray lithography. The PMMA requires a smooth surface to maintain a small and uniform gap between it and the x-ray mask. A small uniform gap allows a better transfer of the mask features to the PMMA by limiting the "fanout" of the x-rays due to diffraction at the absorber edge and/or divergence of the x-ray beam due to less than perfect collimation. Diamond machining is particularly useful with PMMA because of its relatively low glass-transition temperature (105°C). As the PMMA is machined, the heat generated at the cutting edge of the diamond softens or melts the PMMA and tends to recast this soft material behind the tool, resulting in a very smooth surface.

The theoretical peak-to-valley roughness of a surface machined with a round-nosed cutting tool is

$$R_t = f^2/8R \quad ,$$

where f is the tool feed per revolution and R is the tool nose radius in consistent units. The average roughness is

$$R_a = f^2/28.8R \quad .$$

It is easily seen that a smoother surface can be obtained by first reducing the feed but also by enlarging the tool nose radius. These relationships assume that all cutting takes place at the tip of the cutting tool, which is normal for finishing cuts, and the feed is much less than the nose radius of the diamond cutting tool. These relationships are readily found in references on the fundamental mechanics of machining [54,55].

In practice, the smoothness predicted by these relationships is more difficult to achieve because of vibration and deflection of the cutting apparatus, removal of chips without scratching the surface, and the elasticity of the PMMA, which can cause it to deflect under the cutting edge without being removed. In any event, diamond-machined PMMA can exhibit average roughness well into the submicrometer range and even into the nanometer range. This is useful when PMMA resist is cast in a single thick layer, precast sheets are used, or the thickness is created by a series of successive spin-coating operations.

6.2.3 Microdrilling

Microdrilling is characterized not only by small drills but also a method for precise rotation of the microdrill. The walls of a microdrilled hole can also be among the smoothest surfaces produced by conventional processes, next to diamond machining. The smallest microdrills are of the spade type with a geometry as shown in Fig. 6.19. The drills do not have helical flutes as do conventional drills and this makes chip removal more difficult. Microdrills are typically made of either cobalt steel or micrograin tungsten carbide. The steel drills are less expensive and easier to grind but are not as hard or strong as the tungsten carbide drills. The drill point

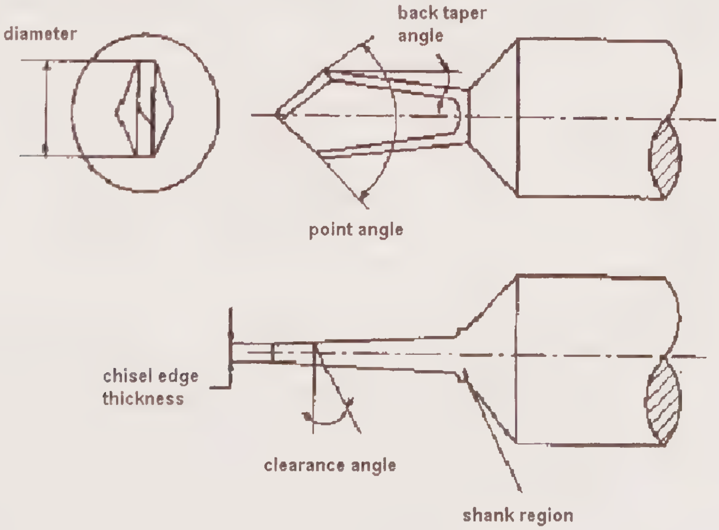


FIG. 6.19 Geometry of microspade drills.

angle is based on the material to be drilled. The normal point angle is 118 degrees, and 135 degress is used for hard materials. The greater point angle provides more strength at the drill point. Although the word “point” is used, drills do not have a true conical point. Instead, the two cutting edges with flank relief meet along a straight, flat cutting edge called the chisel edge. The cutting action along this edge takes place with very high cutting forces, and at the center of rotation of the drill the work material is plastically extruded out from under the drill. For hard materials, this represents a very severe machining condition and one that can lead to holes, with a centerline being offset from the axis of rotation of the drill. The most common failure mode for the drills, past normal handling, is torsional fatigue. This is because as the drilling takes place, a torque builds in the drill as the chip is being cut and when the chip breaks away, that torsional stress is relieved. This results in high frequency, fluctuating torsional stresses.

A microdrilling spindle uses a vee-block bearing arrangement as shown in Fig. 6.20. The drill is mounted in the mandrel and is fabricated integral with the mandrel. The mandrel rides against four convex diamond surfaces, which are the only points of contact. As long as the drill is ground with the mandrel supported in a similar manner, the drill will be concentric about an axis. That axis may not coincide with the mandrel axis but that is not significant as long as the offset is not sufficient to cause excessive vibration, and it normally is not. A small pulley is fastened to the drill mandrel and a drive belt passes around the pulley and drives the drill from an external motor. The belt tension is the only force holding the mandrel against the

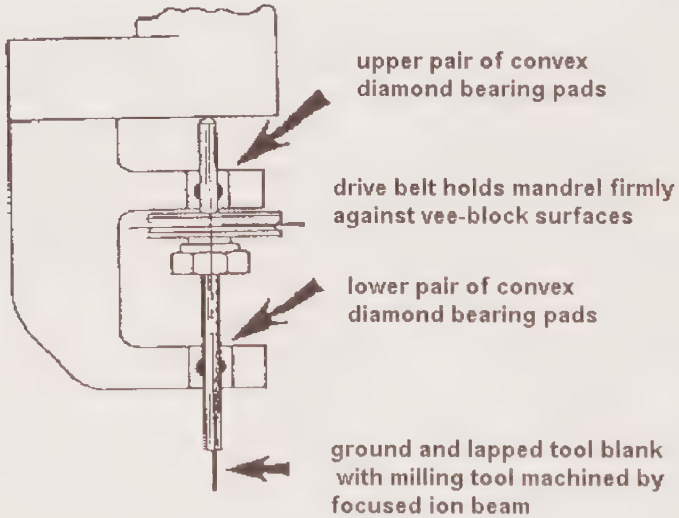

upper pair of convex
diamond bearing pads
drive belt holds mandrel firmly
against vee-block surfaces
lower pair of convex
diamond bearing pads
ground and lapped tool blank
with milling tool machined by
focused ion beam

FIG. 6.20 Vee-block structure used to ensure concentricity of microdrills and micromills during fabrication and use (courtesy National Jet Co.).

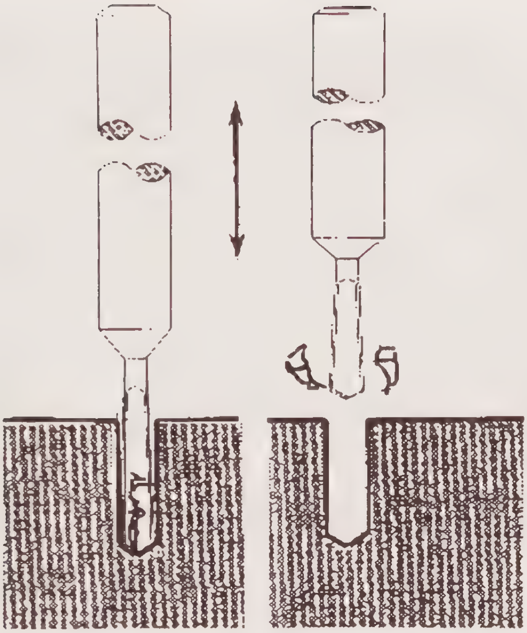


FIG. 6.21 Microdrilling must adopt the peck-drilling cycle to help clear chips from high aspect ratio holes. Thin, flowing fluids (oil or air) increase ship clearing.

diamond pads, and a slight upward component of the belt tension is used to retract the drill. The upper end of the mandrel rides against a ceramic material which provides the drill thrust force. This disk may also rest against a force sensor to measure drill thrust force, which is often used to indicate the extent of drill wear.

Microdrills must be used in a peck cycle in which the drill is repeatedly withdrawn and reinserted into the hole being drilled, as shown in Fig. 6.21. This is necessary to help clear chips from within the hole. A thin cutting fluid is also recommended to aid in clearing chips. The fluid should be moving, as in an air-oil mist, rather than being stagnant. Stagnant fluid will allow chips to reenter the hole along with the drill. The effect of not using a fluid is clearly seen in Fig. 6.22. The hole is completely packed with small chips, on the order of 5 μm in size, and the drilling thrust force under such a condition is typically 3–5 times higher than if the chips are cleared from the hole as shown in Fig. 6.23 [56].

In very soft materials, complete removal of the drill from the hole each peck cycle can cause a slight taper near the hole entrance due to repeated abrasion. This is especially pronounced if the hole centerline is not aligned with the centerline of the drill axis through poor centering of the drill at hole startup. This can be avoided by partial removal of the drill. For softer materials, chip removal is not as severe a problem since the machining forces for such materials are normally lower than for hard materials.

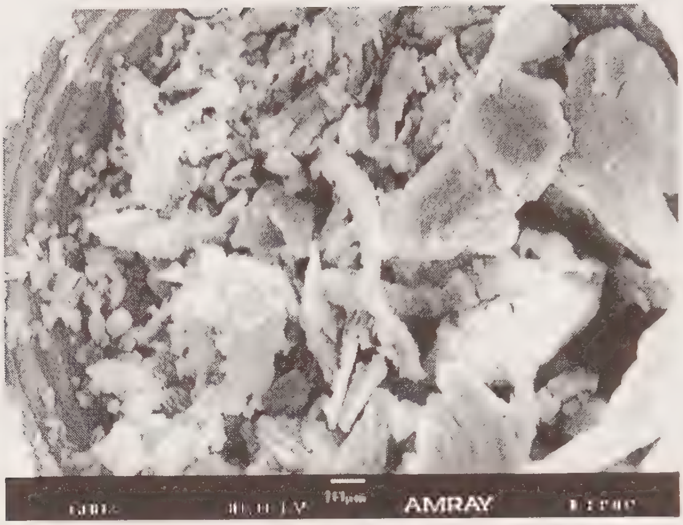


FIG. 6.22 Large amount of residual chips remaining in microdrilled hole in A1-6061 because of lack of cutting fluid.

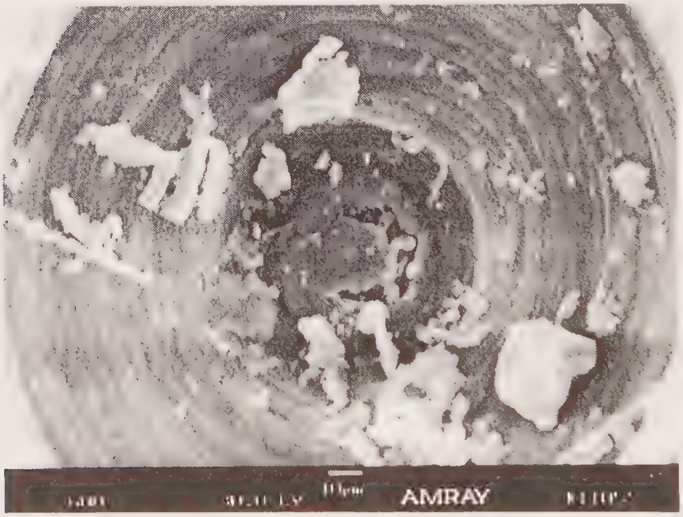


FIG. 6.23 Effectiveness of cutting fluid on removal of chips during microdrilling A1-6061.

The recommended speeds and feeds for microdrilling are as varied as the materials that can be drilled. Microdrilling is not generally a high rotational speed process since dwelling of the drill at the bottom of the hole can cause hardening of the workpiece, leading to increased drilling forces. For most metals, typical spindle speeds are in the 2,000 to 4,000 rpm range and feeds are in the range of a micrometer per revolution or so. Care must be taken when drilling plastics to avoid melting the material, which can lead to adhesion of the plastic to the drill. Melting is caused by high surface speed. This can cause drill breakage or poor sidewall smoothness.

The applicability of microdrilling, as a complementary process, to lithographically produced features has been investigated. A cross section of a copper microgear made by lithography is shown in Fig. 6.24. The average roughness of the hub wall is 0.4 μm. A microdrilled hole in the same material gave a roughness of 0.15 μm over a much longer bore length. Microdrilling can also be used to augment lithography for mesoscopic (millimeter and larger)-sized components. Often parallelism of deep holes is of concern. To determine typical values for paralleleism of microdrilled holes, glass fibers were inserted in a number of holes drilled with a very slow starting sequence. This is necessary to ensure the drill does not walk on the surface of the part and that the hole axis aligns with the undeflected axis of rotation of the drill. Holes with a length-to-diameter ratio of 8 were drilled at 4,000 rpm and a feed of 0.5 to 1.0 μm per revolution. The three-dimensional misalignment of the inserted fibers was measured using scanning electron microscopy to be 0.08 degrees (1.5 mrad), which included skewing of the fiber in the hole due to oversize of the hole which was estimated to be 0.5 μm.

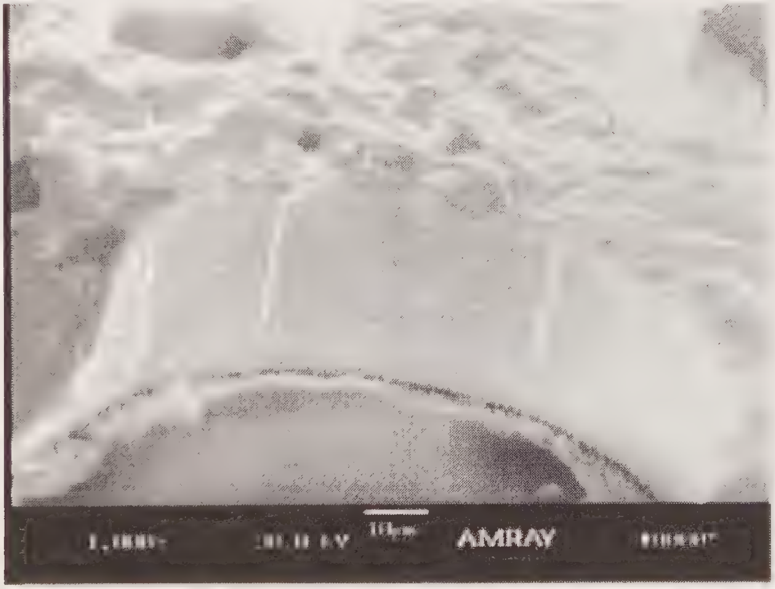


FIG. 6.24 Hub of copper gear made with photosensitive polyimide process.

Microdrilling has one major disadvantage because of the drill geometry. Because of the drill point, a flat-bottomed hole cannot be produced. If one is attempting to produce cylindrical cavities in a micromold, there must be a relatively thick plating base under the mold material, or the structural substrate of the mold could act as the plating base. Full development of the diameter of a hole projected onto a plane perpendicular to the drilling direction requires the drill point to extend 30% of the drill diameter beyond the depth of the fully developed hole. Holes in the 100-μm diameter region require a thick plating base to be deposited. One method for creating flat-bottomed blind holes is to use an end milling tool instead of a drill. The disadvantage of this procedure is that commercial drills have a typical L/D ratio of 4 to 14 while commercial end mills typically have a ratio of 2 to 3.

6.2.4 Micromilling

In conventional machining, no other process is quite as flexible as milling for producing serpentine, two-dimensional or fully three-dimensional shapes in workpieces. For this reason, micromilling has been adapted for the production of molds, masks, and microstructures. This process requires specialized equipment for the production of the milling tools and the actual machining process.

Micromilling tools have been made by the focused ion beam machining process. The micromilling tools begin as a cylindrical blank of high-speed steel (65-68Rc hardness), tungsten carbide, or any other desired tool material. The blank diameter will be the finished tool size and the blank must be created by the same process used to fabricate microdrill blanks. This ensures concentricity of the micromill when used in a vee-block spindle. When used in a collet, the tool must be characterized, which will be described later.

The cylindrical blank then has two diametrically opposed flats machined into it. Because of the Gaussian energy distribution of the ion beam machining process, the blank is turned by 180 degrees after the first flat is machined. This will result in very sharp cutting edges because the portion of the flat toward the incoming ion beam will act as a physical cutoff filter for the normally distributed ion beam. The trailing portion of the flat will be machined by ions which are travelling parallel to the flat surface and will be given a very sharp cutting edge. The edge nearer the ion source will be well rounded and have no cutting function. Therefore, the tool is fabricated for a specified rotational direction, with a clockwise direction looking into the workpiece being standard. A typical micromilling tool is shown in Fig. 6.25. This particular tool has two diametrically opposed cutting edges which have a radius of between 50 and 100 nm as measured with a scanning electron microscope. This particular tool also has the capacity to mill features 70 μm deep because of the long flute. Milling tools and specially shaped electrodes have also been made [57,58] by electrical discharge micromachining, which can be a faster process than focused beam machining but is not as accurate in terms of material removal.

Micromilling should be carried out at a reasonably high rotational speed on a high-precision milling machine. Such a machine is shown in Fig. 6.26. This machine tool

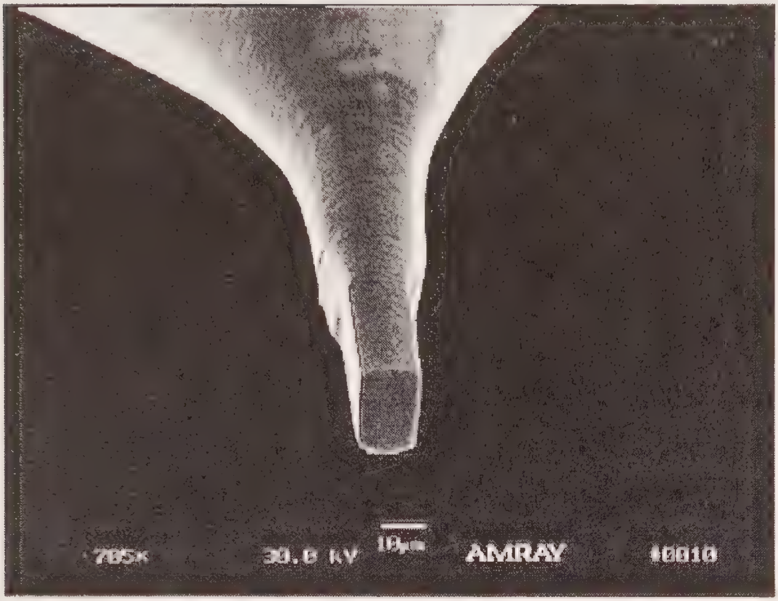


FIG. 6.25 Two-fluted high-speed steel micromilling tool with cutting edge radii approximately 50 nm made with focused ion beam machining.

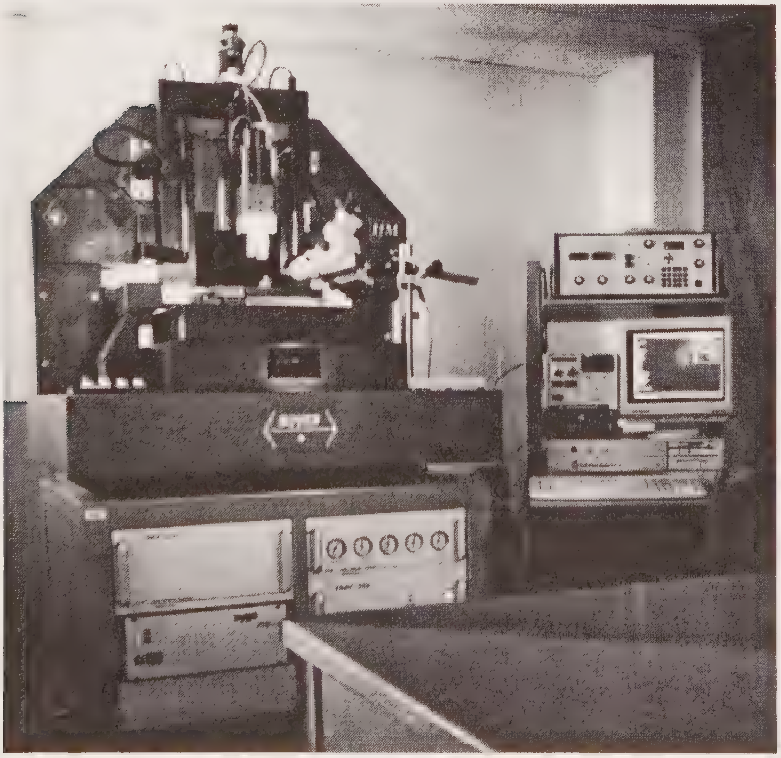


FIG. 6.26 High-precision micromilling machine with laser interferometer feedback on the work table.

has air bearing movements throughout and laser interferometric feedback on the workpiece positioning movements. The milling spindle is capable of turning the tool at 38,000 rpm. Although this is not considered high speed by many conventional machining standards, excellent results have been produced at speeds of 19,000 rpm in polymethyl methacrylate and at lower speeds in x-ray absorbant metals.

Scanning electron micrographs of typical micromilled structures are shown in the following three figures [59]. The trench structures shown are approximately 62 μm deep and 23 μm wide and were produced by the tool previously shown. The trench in Fig. 6.27 illustrates the uniformity, straightness, and vertical side walls possible by this process. The micromilling process also has the capability to produce structures with nonuniform vertical features. The trench shown in Fig. 6.28 was produced by varying the lateral position of the milling tool as a function of depth. In this case, the variation was exponential. Narrow structures can also be milled as shown in Fig. 6.29. The thinnest portion of the wall shown is less than 10 μm wide and 62 μm deep. The entire pattern that contains these individual trench structures is 1.5 mm across and contains three exponential spirals, each covering three revolutions. The pattern typically takes 3 hr to machine to a depth of 62 μm. A typical x-ray mask would have an absorber 10 to 20 μm. A spiral pattern of this depth could be milled in less than 1 hr. This demonstrates the utility of this process for rapid production of high aspect ratio structures.

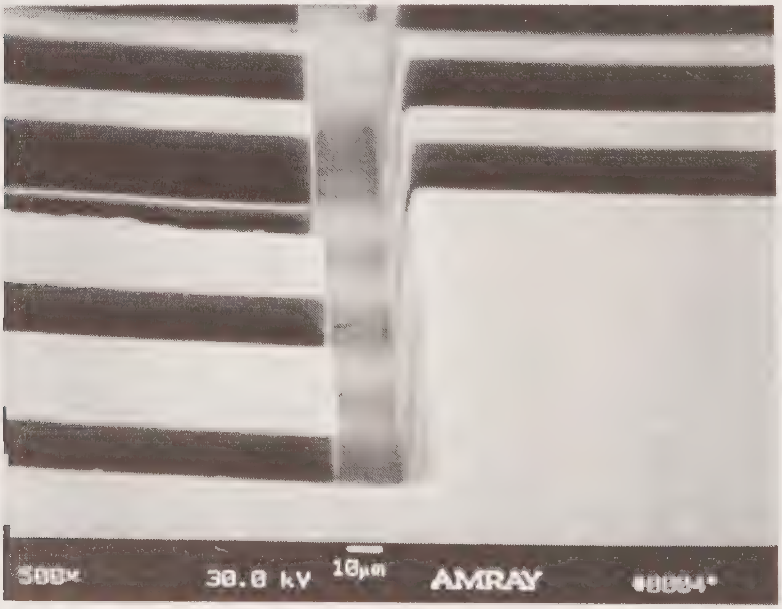


FIG. 6.27 Micromilled trenches with vertical side walls and sharp exterior corners, in PMMA [59]. Trenches are 62 μm deep, 22 μm wide and exterior corner radius is approximately 1 μm.

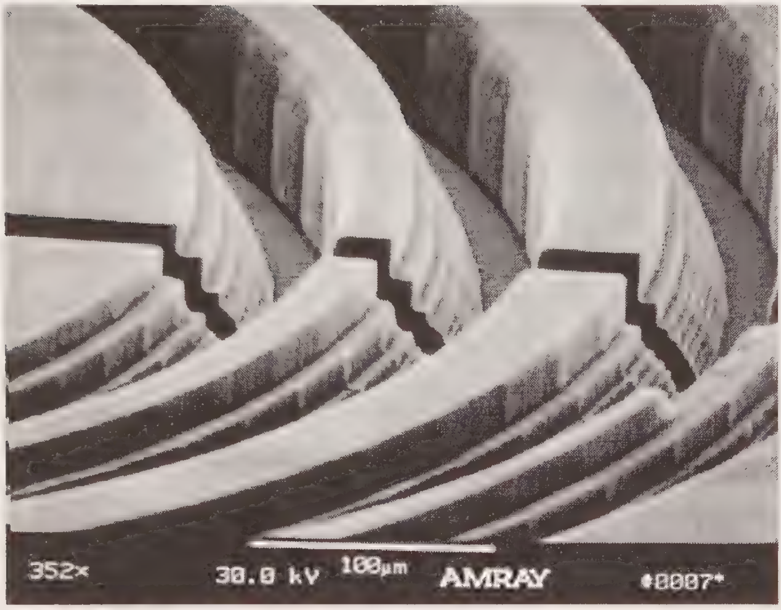

352×
30.0 kV
100µm
AMRAY
#0007*

FIG. 6.28 Micromilled stepped structures in PMMA show flexibility of the process [59]. Each of 15 steps is 4 µm high and narrowest step at bottom is 2 µm wide.

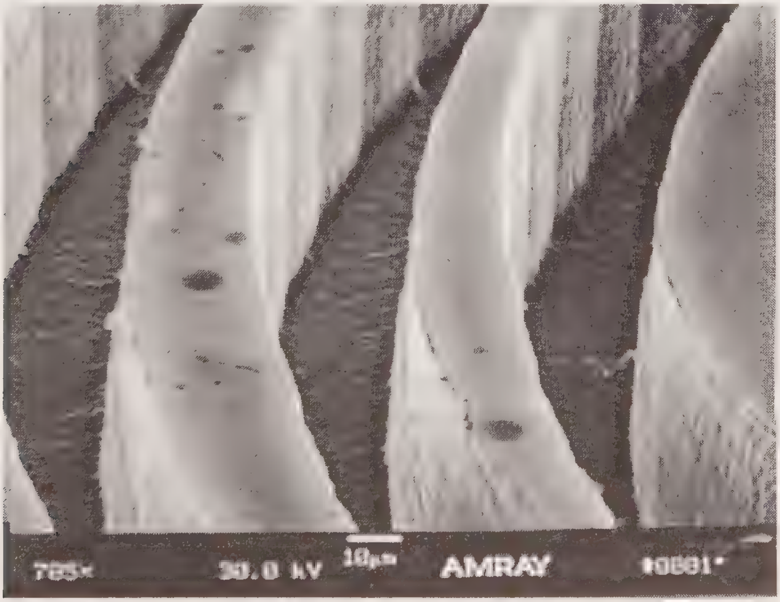

785×
30.0 kV
10µm
AMRAY
#0001*

FIG. 6.29 Micromilled PMMA showing that thin walls are possible. Thinnest wall section is 8 µm wide and 62 µm deep [59].

Micromilling also has the potential to be utilized instead of lithography in low-volume applications. Such an application with the potential for high-volume production is fabrication of x-ray lithography masks. X-ray masks must have materials that transmit and block x-rays. Structural materials with a low atomic number are required for the substrate and high atomic number, machinable materials are required for the absorber. Graphite provides adequate x-ray transmission in thicknesses appropriate for handling and machining. For a 1.3-GeV synchrotron operating with 0.75 nm x-rays, 125- to 250-μm-thick graphite provides adequate transmission to expose 300- to 500-μm-thick PMMA in about 8 hr [60].

Gold absorber is very difficult to machine in a burr-free state. Burrs extending into the opened absorber pattern will detract from the quality of the exposed resist. Likewise, a too thin layer of gold will provide insufficient contrast, as shown in Fig. 6.30. To reduce burring and to increase absorption, alloys of gold with other harder metals such as titanium, tungsten, or tantalum may provide reduced burring. It is also possible to reduce burring by sputtering layers of these materials over gold. A composite mask, such as this, can be more readily machined with fewer burrs. Examples of geometries exposed through a gold-titanium layered mask are shown in Figs. 6.31–6.33. Each of these was machined with 100-μm-diameter commercially available milling tools and the high-precision milling machine previously shown. Micromilling tools are commercially available down to a diameter of 50 μm, and absorber linewidths less than 10 μm wide can be machined with these tools.

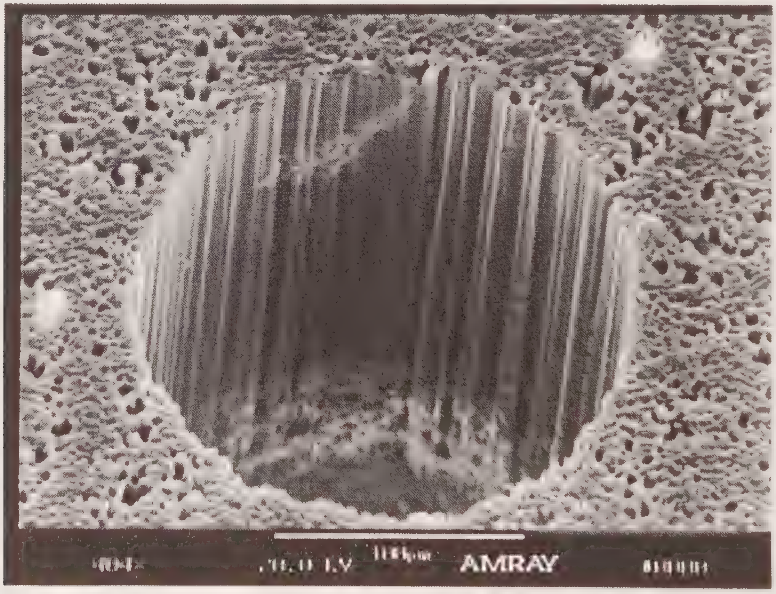


FIG. 6.30 X-ray exposed PMMA through micromechanically machined mask with insufficient absorber to provide good contrast.

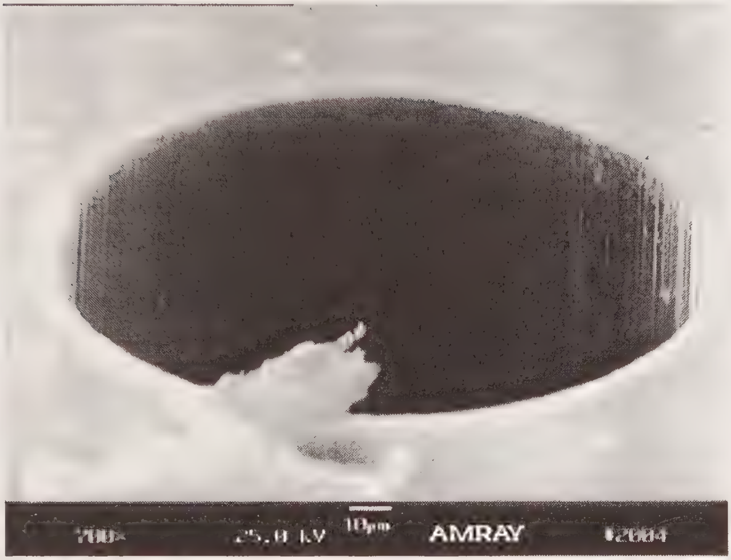

700×
25.0 kV
10µm
AMRAY
#2004

FIG. 6.31 High aspect ratio cylindrical mold created in PMMA by x-ray exposure through a micromilled gold-titanium absorber mask with sufficient contrast. Wall roughness is approximately 1 µm peak-to-valley.

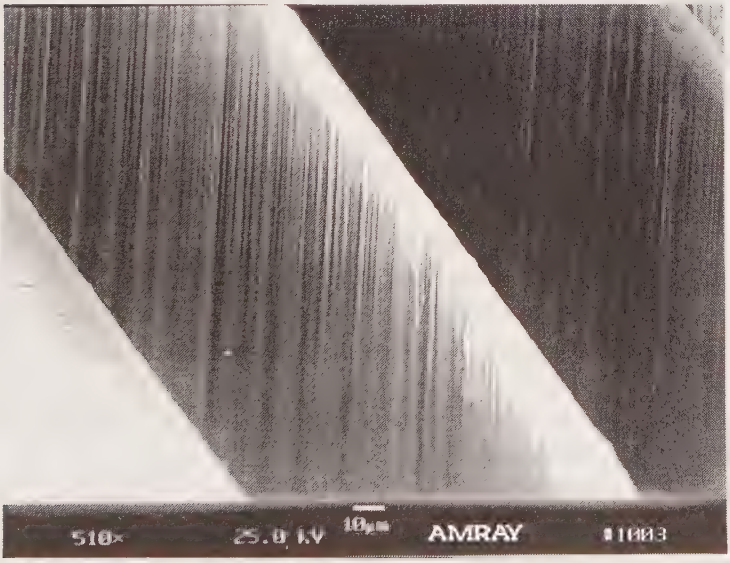

510×
25.0 kV
10µm
AMRAY
#1003

FIG. 6.32 Ten-micrometer wide by 300 µm high wall created in PMMA by x-ray exposure through a micromilled gold-titanium absorber mask. Wall average roughness is approximately 0.2 µm.

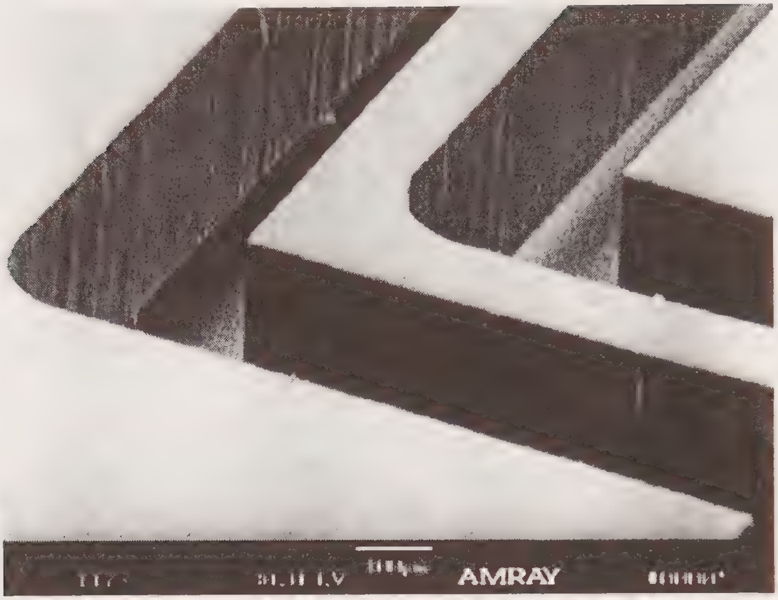


FIG. 6.33 Portion of Joule-Thomson cryocooler mold in PMMA created by x-ray exposure through a gold-titanium absorber mask. Structure is 300 μm thick.

To produce mask or mold structures for actual device applications, the accuracy of the milling process must be addressed. If the micromilling tool is turned in a vee-block, the actual diameter of the tool will produce the machined features, assuming the radial runout is negligible. If a milling tool is held in a precision collet, the machining error is composed of both radial runout and the actual diameter of the tool. In either case, the effective diameter of the tool when cutting must be determined. This can be a relatively simple process if metrology capabilities are present.

The first step is to eliminate, as much as possible, the radial runout. Using a vee-block-mounted tool eliminates this step. If a collet-mounted tool is used, the runout can be nearly eliminated by using a video microscope system or a touch probe to determine the location of the tool cutting edges. The tool is manually "tapped in" to center using a trial and error approach, coupled with the feedback (visual or electronic) from the microscope or touch probe. The runout can normally be reduced to about 3 μm in just a few minutes. After the runout has been reduced as much as practical, the tool must be used to machine two parallel trenches that are closely spaced. Depending on the metrology available, the remaining wall between the two trenches should be of such a size that good resolution of measurment is obtained. By subtracting the measured wall width from the center-to-center spacing of the trenches, the effective cutting diameter of the tool can be obtained. This is then used as a process bias for further machining. Using this technique, the dimensional error in the micromilling process due to tool eccentricity can be reduced to about 0.5 μm [61,62].

6.3 PLASTIC MOLDING PROCESSES FOR MANUFACTURING LIGA AND LIGA-LIKE MICROSTRUCTURES

6.3.1 Introduction

Products of microsystems technology and microelectromechanical systems (MEMS) are considered by representatives of industry to show promise in a future market. However, there are no reliable forecasts of the development and scope of that markets. MEMS are purchased not primarily because of their small size, but because users need components that meet their requirements; they are manufactured reproducibly by low-cost techniques; and they are available in sufficient numbers. Most of the techniques currently used to make MEMS were developed in the semiconductor industry and are based on silicon technology. The consistent application of batch processes has allowed more and more microcomponents (circuits, sensors, etc.) to be manufactured cost effectively on one silicon wafer. Technical bottlenecks hampering a truly widespread use of MEMS are still considered to be the mounting and joining techniques or the sophisticated tests, while the manufacturing processes are deemed to be less problematic.

In contrast, other areas of microsystems engineering—from micro-optics to micro-process technology, to microanalysis, to medical technology, to consumer goods —require an economic process technology to be developed for components made of various materials. This is where plastics molding can make a decisive contribution to volume production. The microstructured tools used for this purpose, which are called mold inserts (in plastics technology), can be manufactured by various microstructuring techniques (mechanical ultraprecision technique, laser technology, reactive ion etching, lithography by means of light or particle beams). The mold inserts are integrated into macroscopic tools. Since the tooling costs are spread over many parts of the molding cycle and over large numbers of units, microstructures can be molded at very attractive prices.

One of the methods of structuring mold inserts is the LIGA technique invented at the Forschungszentrum Karlsruhe (FZK) and used in making microstructures with lateral dimensions on the order of several micrometers, and with high aspect ratios [63–65]. The FZK and other development laboratories in Germany [66], Europe, the United States, and Japan [67] are involved in advancing the LIGA technique by conducting the necessary basic and preindustrial research and then manufacturing microcomponents and microsystems and transferring to industry the product ideas and manufacturing techniques so developed [68–70].

The advantage in molding by the LIGA technique lies in the microstructures being molded out of plastics [71, 72], metals [73, 74], and ceramics in nearly any lateral geometry and with high aspect ratios, with the walls, which are normally perpendicular and plane parallel, having a roughness of less than 50 nm. In this way, even sophisticated structures (high structures, freestanding single columns, oblong and thin walls) can be molded and released from the mold reliably with a relatively broad range of materials.

6.3.2 Molding Microstructures from Plastics

Plastics molding techniques are a precondition for the economic production of microsystems components in large numbers. The processes particularly employed are injection molding, with relatively short cycle times, the rather simple and highly flexible vacuum embossing technique, and for special applications, reaction molding. Molding processes in microtechnology impose special requirements over and above those applied in conventional manufacturing, as will be briefly outlined.

6.3.2.1 Filling the mold

Compact disks (CDs) are among the most sophisticated molded parts in classical injection molding because of the required high surface quality and flatness of the disk-shaped injection-molded part. Molding LIGA and LIGA-like microstructures by injection molding, hot embossing, or reaction molding differs from that technique very clearly in a number of points.

The microstructures in the LIGA mold inserts correspond to blind holes in the mold. When the front of the mold insert is wetted by the molding compound, the microstructures are separated from the remaining mold cavity and, consequently, can no longer be molded. This is why LIGA microstructures require a mold that can be evacuated.

The high aspect ratio of LIGA microstructures implies a high ratio of flow length to wall thickness as well as a high ratio of surface to volume for the molding mass. This implies a rapid heat exchange between the molding compound and the mold. The molding compound, over its entire cross section, assumes the temperature of the molding insert within milliseconds. As the molding compound cools down, its viscosity rises by several orders of magnitude until, even at very high pressure, the cooled molding compound is no longer fluid. Consequently, when molding LIGA microstructures, the mold inserts must be raised to a temperature above the melting point of plastics [75]. In addition, the low viscosity of the molding compounds reduces the mechanical cyclic load acting on the microstructures in the mold insert, thus increasing the service life of that insert.

6.3.2.2 Releasing the mold

After the mold has been filled, the mold inserts and the molding compound are cooled. Thermal shrinkage is offset by the corresponding holding pressure. The demolding temperature is set so that the plastic has sufficient mechanical tensile strength and shear strength. This is necessary for demolding without any flaws, since LIGA microstructures have plane parallel structural walls without any inclination for demolding. One advantage in this respect is the excellent surface quality of LIGA microstructures with mean peak-to-valley heights of R_z=40 nm [76]. This prevents the molding compound from being locked in the structural walls of the mold insert.

6.3.2.3 Cycle time

One molding cycle in injection molding or hot embossing of LIGA microstructures of thermoplastics consists of these process steps:

1. Evacuation and heating of the mold with the microstructured mold inserts;
2. Heating the molding compound to injection temperature and molding temperature, respectively;
3. Filling the mold insert with the molding compound;
4. Applying holding pressure and cooling the mold to the demolding temperature;
5. Venting and opening the mold, releasing the microstructure (either manually or by means of some handling tool).

Cycle times in molding LIGA microstructures are a function mainly of the heating and cooling cycle [71, 75]. Reducing the differential temperature and increasing the heating and cooling capacity allowed cycle times of less than 2 min to be achieved in injection molding plastics microstructures with an aspect ratio below 5 [77, 78].

6.3.2.4 Residual layer

Normally, plastic structures are molded on a substrate of the same material. Microstructures to be molded free from any substrate require innovative demolding techniques. If ejector pins are used, as in classical injection molding, manufacturing and alignment tolerances give rise to gaps and crevices in the mold that are not filled or nearly not filled in classical injection molding. The dimensions of a crevice are the same order of magnitude as those of the desired microstructures. Consequently, in a process of the type necessary for molding LIGA microstructures, they would be filled with molding compound, which might handicap demolding and frequently require reworking.

6.3.2.5 Materials requirements

Compatibility of the molding compounds used with the materials of th mold and the mold insert must be ensured. In addition to oxidation of mold surfaces in air, as a consequence of the high temperatures encountered, there is also corrosion of mold inserts whenever molding compounds contain chemically aggressive components. Corrosion may cause the surfaces of the mold inserts to become rough, which inhibits demolding and which may occasionally damage structural details in the submicrometer range.

These highly specialized applications in micromolding require the use of particularly pure materials specified to narrow tolerances. However, since relatively little material is used even in manufacturing large numbers of microcomponents, no plastics are available on the market which are made to specifications of microsystem technology and qualified accordingly. Consequently, for the time being, standard materials or very small volumes produced in the lab will have to be used.

6.3.3 Tools and Machines for Plastic Molding of Microstructures

6.3.3.1 Mold inserts for plastic molding of microstructures

Plastic molding of microstructures requires special molds. Molds are composed of the mold installed in the machine, which must be specifically adapted to the molding technique, and the molding insert proper as the microstructured part. One characteristic feature of microstructured products is the possibility of manufacturing many parts per mold insert in one molding cycle. The number of parts per cycle can be increased further by integrating several mold inserts in a mold. Especially if only small or medium numbers of one product must be manufactured in a batch, frequent mold changes are necessary. An attractive feature in this respect are molds with replacement units, which allow rapid exchanges of mold inserts. A LIGA mold insert is shown in Fig. 6.34.

In manufacturing mold inserts by the LIGA technique, a resist structure produced by deep x-ray lithography is filled up by electroforming. In contrast to the production of a metal microstructure, the electroforming process in mold insert production is continued until the resist structure has been covered and a complete metal layer is obtained. After removal from the substrate and release of the resist, a microstructured metal part is available as a mold insert.

The thickness of the metal part varies as a function of use. For relatively low micro-

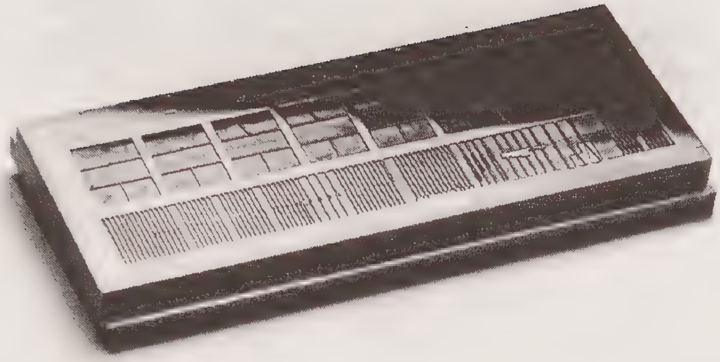


FIG. 6.34 LIGA mold insert manufactured by x-ray lithography and nickel electroforming. Mold insert is 66 mm × 26 mm with 140 μm high microstructures. Smallest features are microprisms with 20-μm-long sides.

structures, such as diffractive elements or in CD manufacturing, 1–2 mm is sufficient (shim technique). In LIGA mold inserts with high microstructures, this thickness is between 7 and 10 mm. The thickness results from the properties of the LIGA technique. Relatively large structured areas give rise to high demolding forces, even in the presence of minor wall roughnesses, which need to be transmitted. Greater thickness not only produces higher mechanical stability of the molds, but also minimizes distortion.

Another special characteristic of microstructured mold inserts is their material. They are made of metals and alloys precipitated by electroforming. This is a type of engineering completely different from classical plastic molding in which the molds are made mechanically out of steel.

The first LIGA mold inserts were made out of a nickel bath of a defined, stable composition because of the low internal stresses attainable in this way. Its relatively high hardness and good chemical stability have justified the use of nickel to this day [74]. Studies of the service life of nickel mold inserts under the scanning electron microscope have shown that no wear phenomena occur, even after 10,000 molding cycles with polycarbonate (PC) [79]. However, certain polymers and polymer modifications, respectively, are known to be incompatible with nickel. Consequently, not every material may be processed with every mold insert. The range of materials has meanwhile been extended beyond electrodeposited pure nickel. Thus, mold inserts can now be made of a variety of alloys [80]. Nevertheless, manufacturing of mold inserts has not yet attained the level of classical tool making.

Many possibilities exist in designing mold inserts. The principles applying to LIGA mold inserts can be compared to those applying to lithographic products: any lateral shape is possible, even oblique edges; two-step or multistep structures can be produced [81]. In addition, LIGA mold inserts have been combined with components made mechanically.

Most mold inserts produced by ultraprecise micromechanics [82] use brass. In manufacturing certain structures, structures produced by micromechanics can also be copied in nickel. Another way of making microstructured mold inserts is by using microstructures etched in silicon. As in the LIGA process, these have excellent surface and sidewall quality [83].

Where the special LIGA properties are not indispensable to the characteristics of the product, such ''LIGA-like'' processes will be preferred for making mold inserts because manufacturing costs are clearly lower than in x-ray lithography with synchrotron radiation.

6.3.3.2 Injection molding

Injection molding of compact disks is a familiar example of injection molding of microstructures. However, as explained earlier, it cannot be extrapolated to injection molding of LIGA microstructures. Consequently, the machine and the process technology used in CD manufacture also cannot be adapted to molding LIGA microstructures.

Injection molding of LIGA microstructures has been under development at the Karlsruhe Research Center since 1986. Major differences relative to conventional injection molding machines are the higher precision achieved to down to the micrometer range—in displacement travel and speeds, the control of the molding cycle specific to microtechnology, and the integration of peripheral equipment. The machine periphery consists of a vacuum unit for evacuating the cavity in the mold to approximately 1 mbar, and temperature control units for the injection mold. It is advantageous to use one temperature control unit each for each half of the mold; this helps to reduce cycle time. Injection molding machines for molding microstructures are available commercially and are operated successfully in a number of research institutions [72, 77, 84] and industries, such as microParts, Dortmund, Germany [70, 75]. All-electric machines for microstructure molding are being developed in cooperation with machine vendors and will be more economic in operation. An injection molding machine for automated production of plastic microparts is shown in Fig. 6.35.

Along with machine development, some mold manufacturers have developed a technique for injection molding of microstructures. The molds can be evacuated. For this reason, they have sealing elements between the mold plates that withstand temperatures of up to 280°C. In addition, the mold plates are permeated by temperature control channels. The temperatures are adapted to the respective half of the mold and the injection and demolding temperatures, respectively, so as to allow a homogeneous temperature distribution in all cavities as well as a relatively short cycle time to be achieved. In addition, the volume of material to be heated in the mold area should be minimized and thermally insulated against the adapter plates [71, 75].

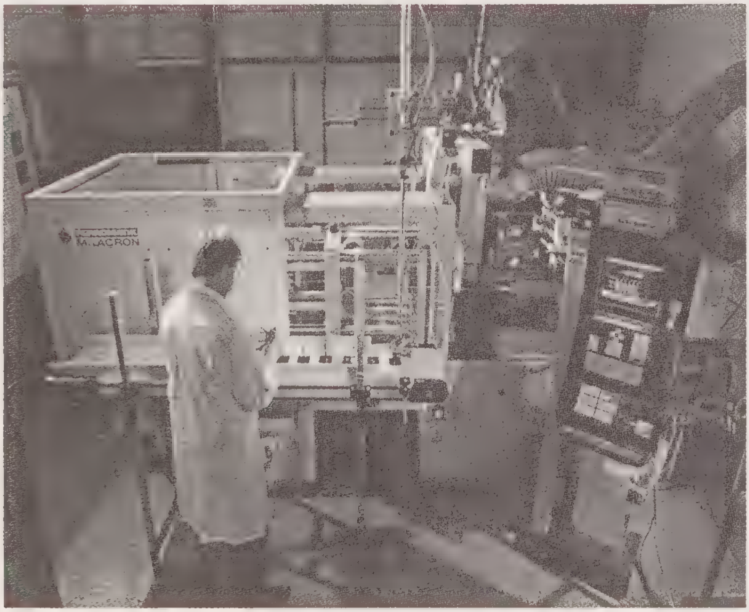


FIG. 6.35 Two-component injection molding machine for the automatic production of microstructures at FZK/IMF III.

At present, microstructures are injection molded of amorphous [PMMA, PC, polysulfone (PSU)], partly crystalline [polyoxymethylene (POM), polyamide (PA), polyetheretherketone (PEEK)], and conductive carbon-filled thermoplastics (PMMA, POM, PA12) with molding compound temperatures up to 450°C in cycle times of 2 to 8 min [77, 84]. The Institut für Materialforschung (IMF III) produces a small series of injection-molded microparts with plastics microstructures which at present have minimum wall thicknesses of 30 μm or structural heights of 1600 μm, with aspect ratios of up to 20.

6.3.3.3 Hot embossing

In hot embossing or thermoforming, the mold insert is pressed into a semifinished thermoplastic product heated to a level beyond the glass transition temperature.

After cooling below the glass transition temperature, the mold insert is separated from the molded microstructures. In principle, this is a very straightforward technique with many flexible uses, which allows some first results to be obtained quickly. For this reason, hot embossing is mostly used for development purposes. As a consequence of the low molding speeds attainable, structures such as freestanding columns or narrow webs can be molded relatively easily. The low material flow allows hot embossing of microparts with low internal stresses, which can be reduced even further by slow cooling.

These characteristics make hot embossing an attractive manufacturing technique also beyond the laboratory scale. Special applications in optics or in prototype manufacture up to the production of small series seem to be possible. For this reason, a concept of an embossing machine for use on an industrial scale has been elaborated [85]. One of the main characteristics of the machine concept is its modular structure. An embossing machine is made up of components available commercially and, on this basis, additions and expansions can be made for complicated microstructures and special applications. An analysis of the requirements to be met in molding microstructures results in these findings: precise control of force, displacement, temperature, and a stable machine structure with precise guidance of the mold so that the product can be released from the mold without being damaged. These requirements are more or less identical with those to be met in the design of testing machines, namely, feed travels on a micrometer scale; force control within a few neutons, with several kilonewtons of load applied; and precise adjustment under external influences. A stable machine frame ensures absolute parallelism of the top and the bottom of the machine.

Control of heating and of the vacuum system can be integrated into the test machine control system (manufactured by Zwick, Ulm, Germany). Over the years, the Institut für Microstrukturtechnik (IMT) has accumulated a high level of competence in designing and building such vacuum embossing chambers. In order to cut costs, an attempt has been made in this component to fall back on as many standard machine tool components as possible. In the meantime, this machine concept has turned out to be so promising that Jenoptik Mikrotechnik, Jena, Germany, agreed to conduct further development and commercialization [86]. Embossing machines can be integrated into a clean room environment as shown in Fig. 6.36.

6.3.3.4 Reaction molding

Reaction molding has been used with materials based on acrylates, amides, and silicones. Because of the low processing viscosity of cast resins, relatively simple and good filling of the mold and excellent molding accuracies are obtained. To support demolding, internal demolding agents are added in some cases.

In principle, the setup of the reaction molding machine at FZK corresponds to that of the hot embossing machine. In addition, the reaction molding machine is equipped with a mixing and metering unit for the cast resins as well as a temperature control unit for heat supply to the resin systems solidifying under hot conditions, and to remove the reaction enthalpies. After reaction hardening of the plastic, demolding is achieved as in the hot embossing process.

FIG. 6.36 Embossing and welding machines making LIGA microspectrometer components in a clean room environment at FZK/IMT.

To make use of the low viscosity of reaction-cast resins for good filling of the mold and, on the other hand, achieve short cycle times, molding of polymer precursors by UV-induced polymerization is under development ("photomolding"). For this purpose, a mold that can be evacuated and is partly transparent has been designed for reaction molding at the appropriate temperatures. UV light is fed to the mold cavity from a powerful light source. Reaction shrinkage is compensated by the appropriate holding pressure, and demolding is achieved by ejection pins [87]. In addition, reaction molding with a cold curing system is used also to make molds out of silicone (polysilazane). The silicone mold is used to produce lost molds by simple wax microstructures for the development of molding techniques for ceramic microstructures.

6.3.4 Molded Microparts

Molding is a key technology in the economic process of manufacturing microparts and microcomponents. It is becoming more and more clearly evident that molded microparts are going to be used in nearly all industries. A few examples are outlined in the following sections.

6.3.4.1 Microfluidics

Pumps and valves are frequently used in process technology and for analytical purposes. Both microcomponents have already been made out of injection-molded housing components. In the micropump [88] developed at the Karlsruhe Research Center, the microstructured housing shells are injection molded out of PSU. The pumps are manufactured in small series in blanks of twelve each by bonding the housing sections together with a polyimide membrane, which carries a helical heater for a thermopneumatic drive. Such a pump is shown in Fig. 6.37.

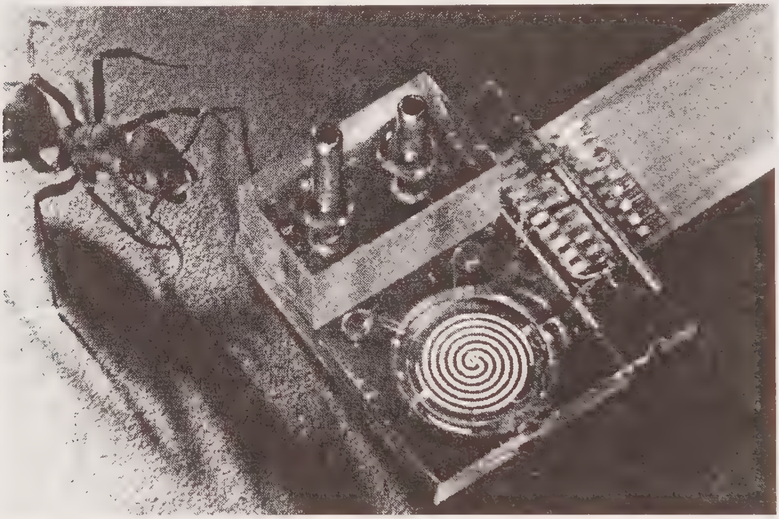


FIG. 6.37 Micropump fabricated at FZK in small quantities, to scale with ant. Pump chamber diameter is 3.8 mm. Typical flow rate is 200 μl/min without back pressure [120].

The same thermopneumatic actuator principle is used in microvalves [89]. At the same time, an electrostatically actuated microvalve was developed. For the electrostatic microvalve, the parts of the housing are injection molded of PA12 filled with conductive carbon, and act as electrodes [90].

6.3.4.2 Micro-optics

The market for micro-optical components seems to promise high growth rates. This is due in particular to increasing uses in telecommunications. It is not so much high-performance components, but components produced under conditions favorable for broad areas of application which are in the foreground of interest and, consequently, constitute an ideal application of plastics molding. Low materials prices, a rich choice of materials, and high production outputs open up many opportunities in this area. The number of optical components to be found in this market segment is correspondingly large [91, 92].

The most commonly used elements are diffractive elements with lateral structures in the nanometer range. The structural height is not much greater (aspect ratio ≈ 1). What matters is the excellent surface quality, which can be achieved by various molds in combination with the hot embossing technique. The main products to be mentioned in this respect are Fresnel lenses [93] and holograms [94].

Slightly higher structures are required to make monomode waveguides. In addition to the optical quality required, these elements also demand strict separation of materials in order to avoid coupling losses [95]. Another important area consists of multimode applications. This is one of the main fields in which plastics techniques are used, and a large number of component developments meet a variety of applications. Important components are connecting structures, such as plugs [72] and Y-couplers [70]. Such components, which are injection molded in the submicrometer range with structural heights in excess of 1000 μm, aspect ratios around 10, and structural details in the submicrometer range, are typical LIGA microcomponents indispensable in the design of optical networks. If various elements of micro-optics are to be combined, such as spherical lenses and optical fibers, micro-optical benches are absolutely essential. Also, these microcomponents, which are typical of the LIGA process, must meet the most stringent requirements of dimensional stability for three-dimensional support structures [91, 96–99].

Hot embossing is the technique of choice for this structure and Fig. 6.38 illustrates the process scheme. In the first step, the waveguide core material is pressed into the mold insert, forming a plane surface with the tool face. In a second step, a polymer sheet made from the cladding material (a copolymer of PMMA and a fluorinated methyl methacrylate) is welded at the glass transition temperature to the core layer by applying sufficient pressure. At the interface, polymer diffusion occurs in a layer 1–10 μm thick, which results in a very strong adhesion of this layer. After demolding, the optical fiber is inserted into the fiber groove and the reflection grating is produced by depositing a metal onto the grating. Finally, a UV-curable polymer is applied onto the structure, which then forms the second cladding layer.

Lenses are other important optical components. Their dimensions are large compared to those of diffractive elements. Again, the requirements to be met in surface quality demand the highest precision. Various approaches exist to achieve the curved lens surface for producing mold inserts. One way is to remelt cylindrical LIGA microcolumns and further process them into mold inserts [38]. Molding by hot embossing in polycarbonate produces microlenses of excellent quality, used in microobjectives because of their imaging accuracy. Similar results can be obtained by molding with a silicon mold insert in which lens-shaped impressions were made by isotropic etching. Surface quality of silicon in this case determines the quality of the plastic lens [100–102]. In both cases the surface quality of the microlenses is on the order of 10 nm–30 nm Ra value and acceptable for optical applications.

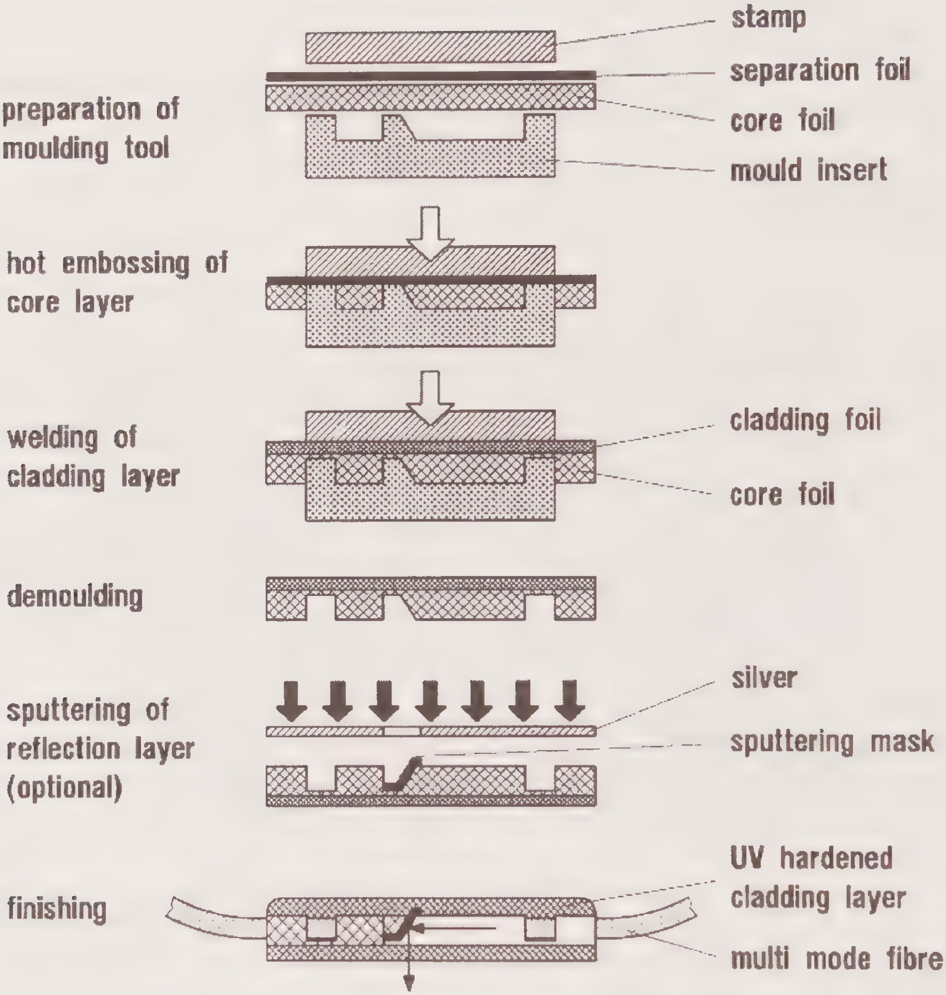

stamp
separation foil
core foil
mould insert
preparation of moulding tool
hot embossing of core layer
welding of cladding layer
cladding foil
core foil
demoulding
sputtering of reflection layer (optional)
silver
sputtering mask
finishing
UV hardened cladding layer
multi mode fibre

FIG. 6.38 Hot embossing process sequence for optical fiber coupler.

One final example is to draw attention to the LIGA microspectrometer (shown in Fig. 6.39), a product that is already commercially available. In principle, this is a waveguide structure, but as far as function is concerned, a microspectrometer is a complete microsystem. In addition to the beam guide system in the waveguide, a self-focusing diffraction grating, as well as a launching shaft and a sloped edge for coupling out onto a diode array, are integrated components. A system of this type never needs to be adjusted; this feature, in addition to production by plastics molding, is the reason for the low price. (Further technical information is available from microParts GmbH, Hauert 7, D - 44227 Dortmund, Germany.) From the point of view of production technology, two types of spectrometers are distinguished. The UV/VIS spectrometer consists of three layers [103], with light being conducted in an embossed core layer. A major problem in this design is the sharp transition from

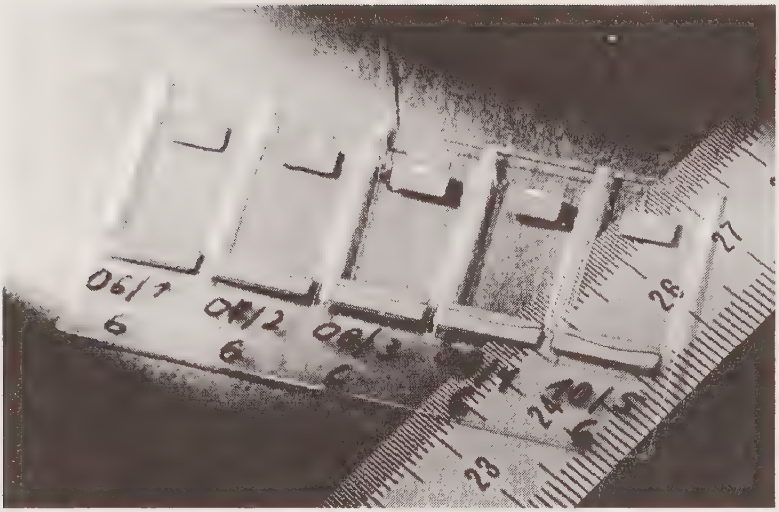


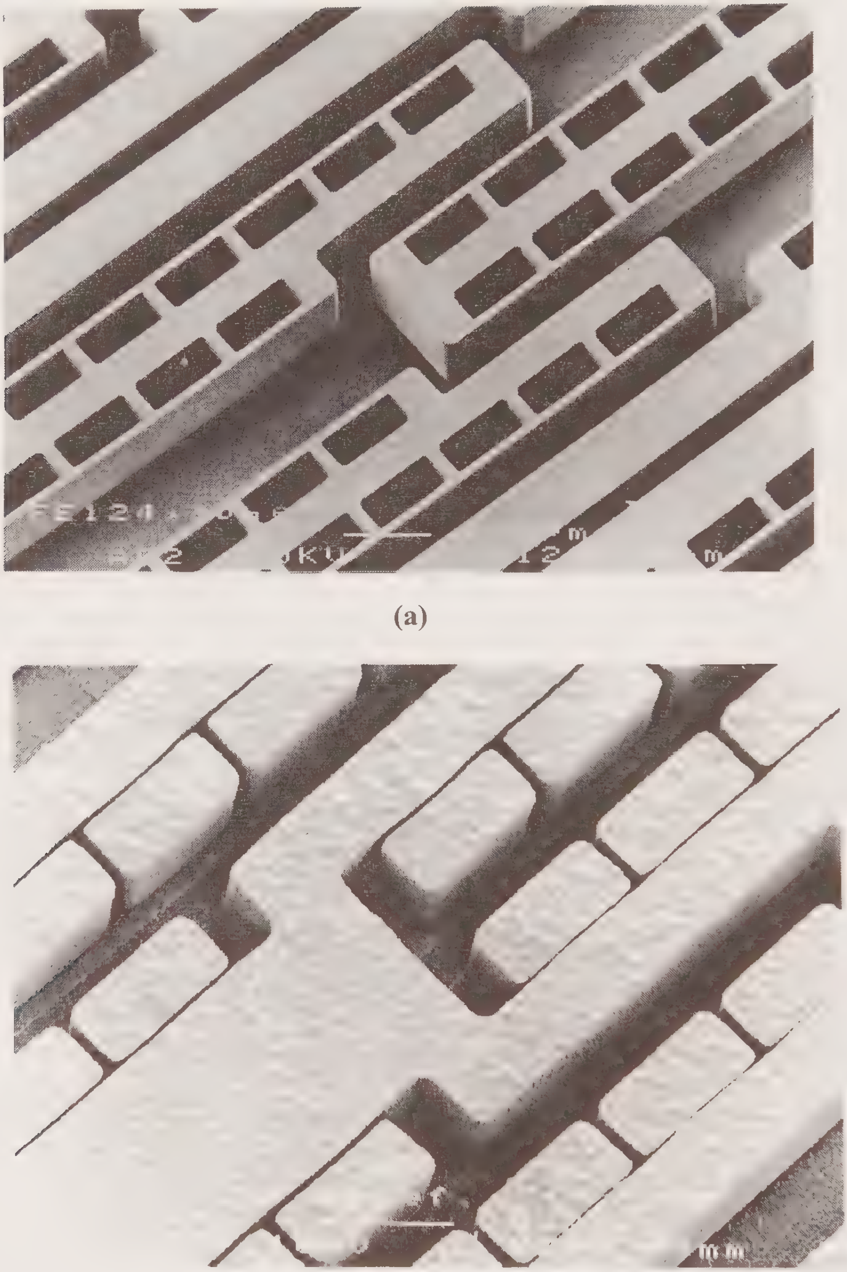
FIG. 6.39 A set of five microspectrometer components manufactured in one batch by hot embossing and welding from PMMA and a fluorinated PMMA copolymer, respectively. The grating and the 45-degree edge were covered with a reflecting metal layer by sputtering.

the structured core layer to the unstructured cladding layer, where loss of light into the cladding layer must be avoided. Hot embossing is the technique of choice in such problem cases.

The infrared spectrometer is a hollow waveguide component in which light is conducted in air. Microstructures are integrated into the support structure and the light conduction structure. The advantage of this technique lies in only one material being employed, allowing large numbers to be produced by injection molding [104].

6.3.4.3 Sensors and actuators

Embossed plastic microstructures are used as lost molds when making movable metal microstructures [34]. This process requires substrates analogous to the familiar sacrificial layer process. Such a structure is shown in Fig. 6.40. Structuring by aligned x-ray depth lithography is replaced by positioned embossing on the prestructured substrate. There are other benefits besides the cost advantage of the molding technique over x-ray lithography. If the substrate is a wafer with the electronics part integrated in the CMOS technique for triggering or signal evaluation of the movable microstructure, structuring is possible only by embossing [105]. The high-energy x-radiation would destroy the sensitive electronic structures. One example of a movable microstructure in the field of sensors is the temperature-compensated LIGA acceleration sensor [106]. This component penetrates into measuring ranges that used to be limited to expensive precision sensors.



(a)

(b)

FIG. 6.40 (a) Plastic microstructure on the sacrificial layer for manufacturing LIGA acceleration sensors, and (b) metal acceleration sensor microstructure manufactured by nickel electroforming using lost plastic mold in (a).

Another device fabricated by the positioned hot embossing process is a microrelay as an example of an actuator. In this device, the movable component, an armature suspended from spring elements, is electroplated from a soft magnetic material (FeNi alloy) [107].

6.3.4.4 Lost micromolds

Lost micromolds manufactured by injection molding (e.g., from PMMA or wax) are used to fabricate ceramic microstructures, such as spinning nozzles or lead-zirconate-titanate (PZT) column arrays. Other processes for manufacturing ceramic microstructures, such as ceramic injection molding (microCIM), are under examination. In addition, ceramic microstructures are used as lost molds for the microcasting process for metal microstructures and experiments for metal injection molding (microMIM) with an iron powder-filled feedstock have been made [73]. The term 'lost micromold' describes the molding of a polymer microstructure by injection molding or hot embossing which is further replicated into other materials like metals using electroplating or ceramics using slip casting. These microstructures form the final product and the intermediate polymer microstructure will be removed and, by doing so, gets lost.

Conductive-filled thermoplastics were investigated for their suitability for further development of injection molding of electroformable plastic microstructures. Thermoplastics with a relatively high content of conductive carbon black, such as different types of POM-C and PA12-C, were selected and an example is shown in Fig. 6.41. To investigate the general capability for electroforming, plate-shaped semifinished products were made with variable injection parameters. Then the

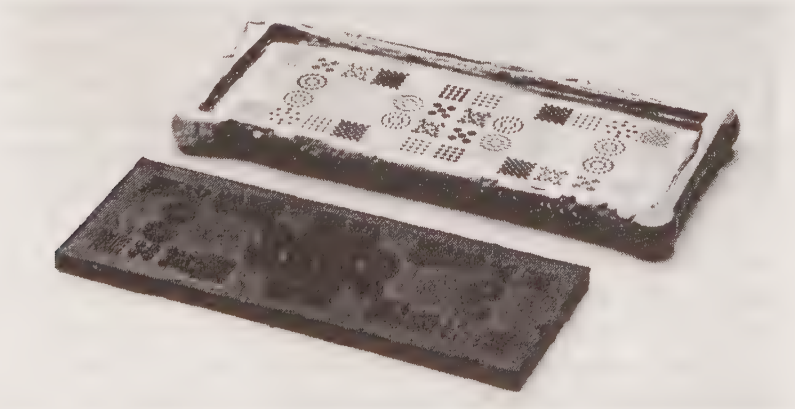


FIG. 6.41 Lost micromold injection-molded part with conductive-filled PA12 (bottom), and the copied electroformed nickel test structures for spinnerette nozzles (top). Mold insert is 66 mm × 26 mm × 1.3 mm high with smallest components being 270 μm diameter.

surface resistance was measured. Injection speed was found to be the main parameter influencing surface resistance. The surface resistance was found to increase with increasing injection rate. Previous experiments showed that a surface resistance of less than 1000 ohms is required for electroforming high-quality microstructures. Hence, it has been ascertained that carbon-filled POM-C and PA12-C are electroformable in case of injection molding in the lower speed range. This has also been confirmed by electroforming tests with semifinished products [71].

Using the molding parameters obtained from injection molding of semi-finished products, lost molds with microstructures of conductive filled PA12-C were manufactured for electroforming spinning nozzle test structures. Both the lost plastic mold and the electroplated spinning nozzles are shown in Fig. 6.41. In addition, free-standing two-stage LIGA tooth wheels with 550 µm of structural height and an aspect ratio of 15 were made by electroforming on PA12-C in nickel [78].

6.4 DEVELOPMENT OF CERAMIC MICROCOMPONENTS

6.4.1 Introduction

The development of modern high-performance ceramics has increased in importance during recent years. This is mainly due to the specific properties of the ceramics, which make them superior to other classes of materials. The most typical of these properties are their high hardness, abrasion resistance, resistance against corrosive media, and high temperature applicability.

In order to realize this potential in the field of microsystem technology, development efforts have to be done in the field of material synthesis as well as processing techniques. Considering a microstructure with overall dimensions of a few millimeters and internal dimensions in the range of microns, it is obvious that there are much higher demands upon the raw material and the processing techniques compared with larger structures. A very interesting class of ceramics for microsystem technologies consists of multicomponent functional ceramics. To maintain special properties and isotropic behavior in even the smallest part of the structure, it is absolutely necessary to achieve chemical homogeneity in the starting powder. Furthermore, in order to obtain good molding results and high form fit of the ceramic part, the average grain size of the sintered part must be at least one order of magnitude less than the minimum internal dimension of the structure. In the following section, an overview is given of processes, taking into account synthesis as well as shaping, which have been developed for the production of microshaped ceramic parts.

6.4.2 Thermal Two-Stage Process for the Production of Multicomponent Ceramic Powders

A special process has been developed for the production of multicomponent oxide ceramic powders. The component materials are metal organic liquid precursors, e.g., alcoxides or acetates. These precursors can be mixed to obtain an ideal solution. In some cases the components react with each other to build complex compounds with alcoholate ligands and with the metals as central atoms, interconnected via oxygen atoms. By spray drying this precursor solution, a precursor powder can be obtained as a first step. In the second step, this powder is converted into the desired oxide ceramic powder by a thermochemical conversion. This process is performed in a high-temperature fluidized bed system. The laboratory fluidized bed system developed at FZK [108] allows thermal processes to take place at temperatures up to 1,200° C, involving gases of any composition and batches of solids up to about 1 kg. Due to the favorable heat transfer of the material moved in the turbulence of the fluidized bed, isothermal powder annealing treatments as well as chemical reactions can be performed. By this process, very homogeneous and well-sinterable powders have been produced. Up to now a great variety of powders have been prepared on a laboratory scale, including lead-zirconate-titanate, doped with different additional elements; lead-metaniobate also with additives; and indium-tin-oxide.

6.4.3 Shaping Techniques

Special shaping processes had to be developed for the production of ceramic microcomponents, because most processes for shaping microcomponents made of metallic or polymer materials are not transferable to ceramics, and standard processes, such as dry pressing of powders, are not applicable. The special shaping processes include:

1. Tape casting and stamping;
2. Slip pressing;
3. Direct electrophoretic sol deposition.

The *tape casting and stamping process* [109, 110] starts with an organic slurry consisting of ceramic powder with organic additives, such as solvents, binders and plasticizers in order to form a stable suspension. This slurry is intensively mixed and then cast to tapes and levelled to thicknesses of about 500 μm. These tapes are then allowed to dry at ambient conditions. After drying, the tapes are cut into even pieces to fit into the stamping device. Stamping is done with metallic dies, which can be fabricated either by micromachining [111] or by electroforming of LIGA structures [112]. The stamping is performed in an axial press with controllable pressure. After stamping, the tool is removed from the ceramic part, which is then debindered and sintered. A schematic flow diagram describing this process as well as the slip pressing method is shown in Fig. 6.42. Up to now, several structures from different ceramic materials have been produced by this process. The minimum internal dimensions, that can be achieved by this process are about 20 μm, with aspect ratios of up to 6, depending on the design of the structure. Examples are shown in Fig. 6.43. The advantage of this process is its simplicity and its potential for large-scale production.

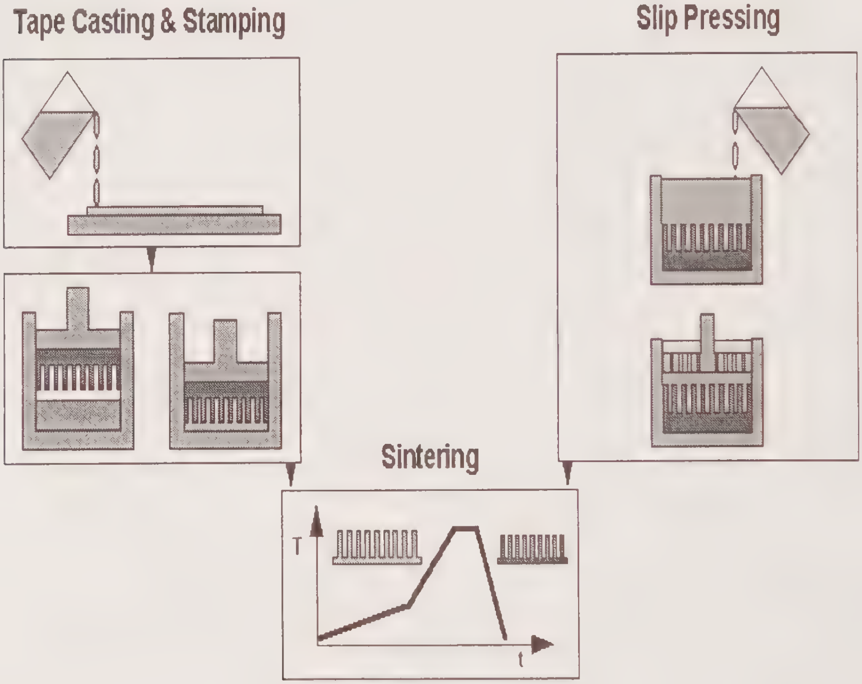


FIG. 6.42 Schematic diagram of the tape casting and slip-pressing molding processes.

The slip pressing process [113, 114] starts with an aqueous slurry and uses lost molds. The mold, being the inverse structure of the desired microcomponent, is fixed at the bottom of a cylinder, into which the slurry is filled. The liquid is then driven out by axial pressing with a porous upper piston. The mold, mainly PMMA, has to be removed from the resultant green body. This can be done either by pyrolysis or by a chemical solvent. After this, the ceramic microcomponent can be sintered. By this process, structures with minimum lateral dimensions of about 20 μm and with thicknesses up to about 1 mm have been formed, which means that aspect ratios of about 50 have been achieved.

Direct electrophoretic sol deposition is a method in which the shaping is done with the sol as liquid precursor without using any powder [113, 115]. No powder conditioning or further treatment is necessary. This process was developed for shaping piezoelectric microcomponents consisting of lead-zirconate-titanate. In this simple process, two electrodes are inserted into the sol in a defined distance and an electric current is applied. While the cathode is a platinum sheet, the anode is the negative of the desired microcomponent. Standard microshaped polymer molds (PMMA) are used that have been sputtered with gold. The sol particles are deposited at the gold surface, gelation occurs simultaneously, and over time, a layer of PZT gel is grown on the PMMA mold.

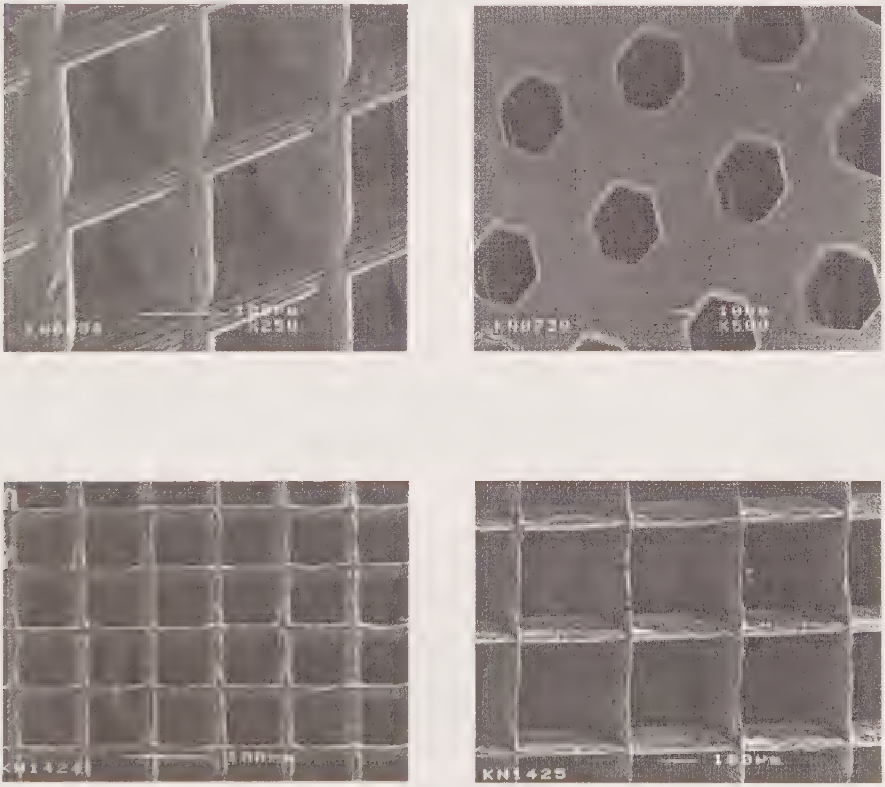


FIG. 6.43 Four examples of ceramic microcomponents molded by tape casting and stamping.

After successful deposition, the PMMA mold with the adherent gel layer is removed from the sol and the gel is dried in order to remove the solvent. The next processing steps are the thermal decomposition of the PMMA mold, the burnout of all organic additives in the gel, the conversion into the ceramic state, and the sintering process. By this process, special PZT shapes have been prepared as shown in Fig. 6.44.

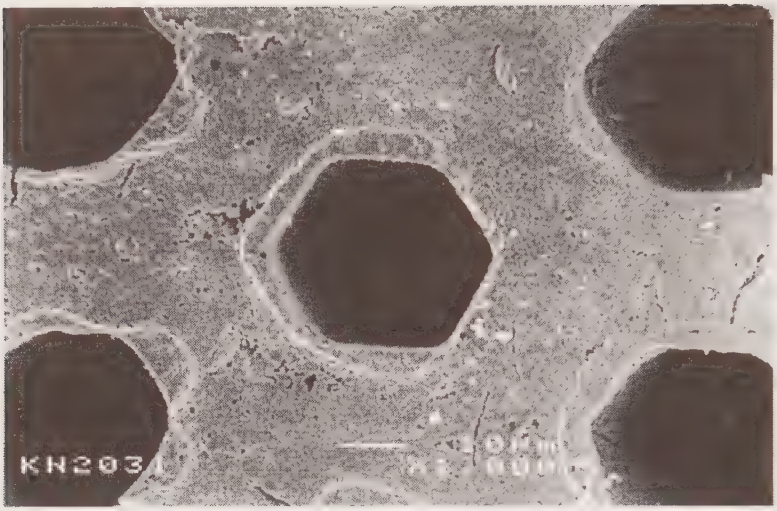


FIG. 6.44 Piezoceramic (PZT) microcomponent molded by direct electrophoretic sol deposition.

6.4.4 Development of Microcomponents

Several ceramic microcomponents have been developed by the processes described earlier. It is possible to produce microshaped ultrasonic transducer arrays from PZT [116] using tape casting and stamping (see Fig. 6.45). The geometry is as follows: The active elements are square-shaped columns with cross sections of 120 by 120 µm and about 350 µm high. The array dimensions are 20 by 20 mm. The width of

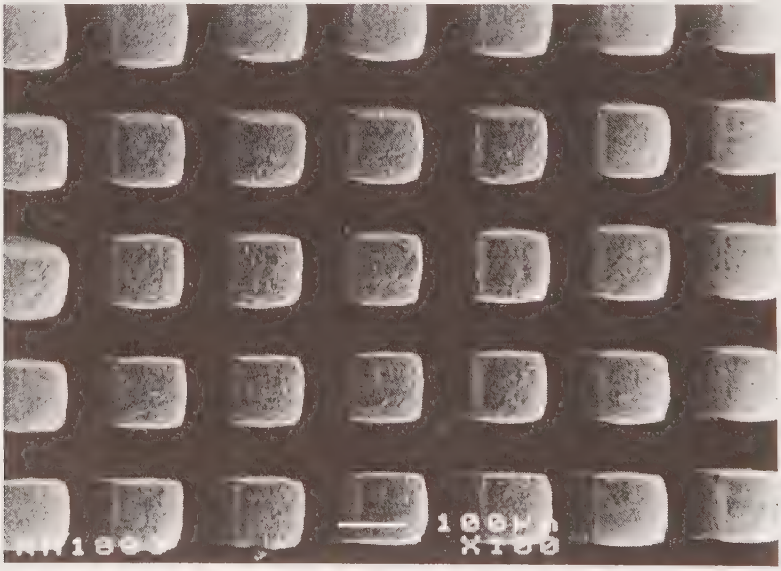


FIG. 6.45 Detailed cutout of the ultrasonic transducer array structure.

the trenches is about 80 μm. After sintering the array, these trenches are filled with polymer and the residual bottom is removed by grinding. The resultant piezoelectric composite is then electroded and poled. The characteristics of this piezo element can be tailored for various transducer applications [117] by varying the geometry of the electrodes and by electronic control.

A microshaped ceramic heat exchanger has been developed using the tape casting and stamping [118] method. First the ceramic tapes are stamped with a metal die showing the inverse structure of the exchange channels. Then these green structures are laminated crosswise, giving the base element for the later crossflow heat exchanger. This element is then debindered and burnt. The external media connections are formed by slip casting and will be connected to the base structure by diffusion bonding or by ceramic soldering. This latter step is still under development. Figure 6.46(a) shows a cross section of the exchanger unit with channel widths of 120 μm. The overall dimension of the exchanger unit is about 1 cm [see Fig. 6.46(b)].

Ceramic nozzle plates have been fabricated by slip pressing [114] as shown in Fig. 6.47. Because of their high wear resistance, there is an increasing interest in the development of ceramic nozzles for the spinning of textile filaments. Various geometries can be verified by this method, which shows a very good quality of replication, even for complex cross sections, which cannot realized by mechanical machining, and for very high aspect ratios.

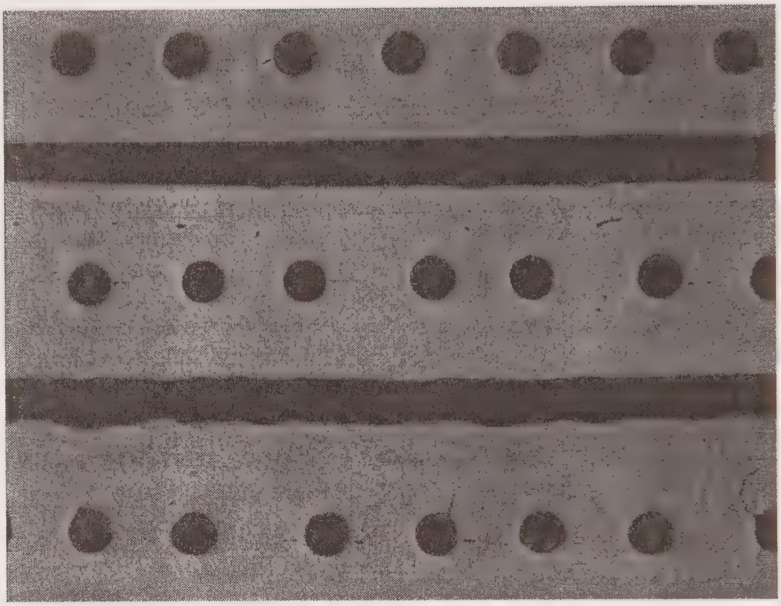
6.4.5 Conclusions

The production of ceramic microcomponents requires development efforts in the field of materials synthesis as well as in the area of shaping techniques. The processes described here are applicable for a wide variety of materials and microcomponents. They provide the basis of a processing technology that can help to solve different problems in the development of ceramic microcomponents.

6.5 SUMMARY

The material contained in this chapter has shown that the LIGA process offers many advantages:

1. High structural accuracy and high aspect ratio due to the patterning qualities of x-ray lithography, which offers significant advantages with respect to hybrid microsystems where maintaining dimensional accuracy makes assembly easier using LIGA components (LIGA leggo).
2. Reproducibility, repeatability, and reliability achieved by the use of high-precision replication molding techniques.



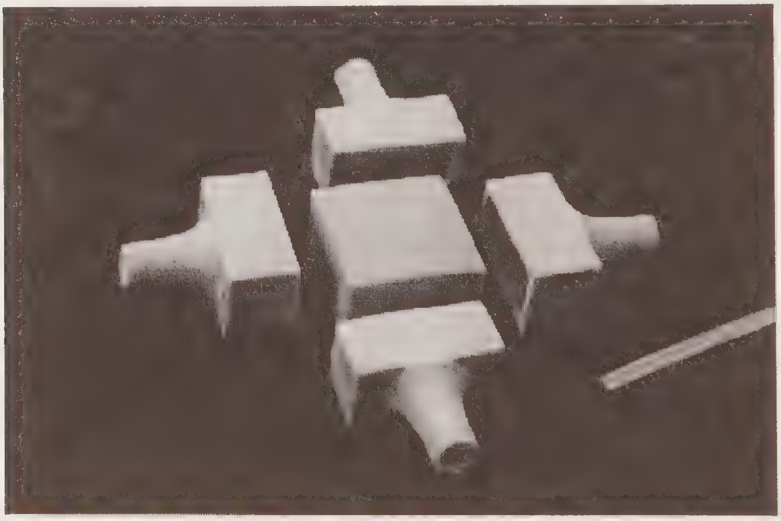


FIG. 6.46 Cross section (top) and overall view (bottom) of the ceramic micro heat exchanger with channel diameters of 120 μm.

3. The capability of using a broad range of materials to address a variety of applications.
4. An inherent flexibility that provides a high level of compatibility when the integration of LIGA processes with other microfabrication technologies is required.
5. A relatively simple system integration at the assembly and packaging levels.

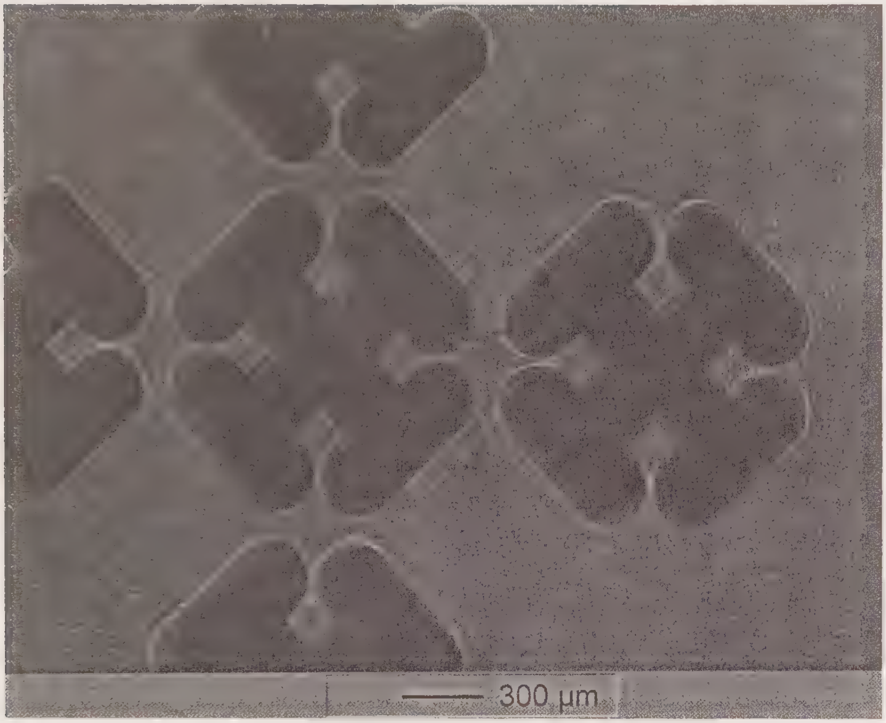


FIG. 6.47 Detail of ceramic (Al_2O_3) spinnerette nozzle.

While micromechanical machining processes cannot provide the same minimum lateral dimensions and level of sophistication as LIGA does, they nevertheless provide a method for the rapid production of molds or final parts in a wider variety of possible materials. Together, the two approaches for creating high aspect ratio structures give the microsystem designer wide latitude in design geometries and process methods.

Many additional resources are now available via the World Wide Web and a representative list is given below.

MEMS Clearinghouse (provides general information)
http://mems.isi.edu/
http://mems.isi.edu/mems/yp/micro.html (MEMS Fabrication/Services Providers for Micromachining)
http://mems.isi.edu/mems/yp/app.html (MEMS Applications, companies manufacturing or providing services related to MEMS Applications)
http://mems.isi.edu/mems/yp/mems_centers/index.html (MEMS Fabrication Capabilities in various institutions and organizations)

European Microsystem Technology Home Page (EMSTO)
http://www.cf.ac.uk/ecad/EMSTO/emsto.html
http://www.msrc.cymru.net/default.htm#pagetop

LIGA-news
http://www.uni-mainz.de/IMM/LNews/LIGA_home.html

Available from the MEMS Clearinghouse WEB page are the following addresses:
Analog Devices i MEMS Project
Analog Devices US East Coast Mirror site for the i MEMS Project
The ARPA/ETO Microelectromechanical Systems
Asian Technology Information Program
CAEMEMS Software WWW site from the Univeristy of Michigan
The Catholic University of Louvain, Belgium
The Center for Space Microelectronics Technology
DIMES Delft University, Delft, The Netherlands
Electronic Instrumentation Laboratory, Delft University
Electronic Subcommittee (ESC) NAVIGATOR
IEEE Micromechanical Systems link
The Institute for Micromanufacturing (IfM) at Louisiana Tech University
JTEC/WTEC Program on MEMS at Loyola College
The MEMS Interchange
MEMS program at Northeastern University
MEMS Research at the University of Wisconsin - Madison
MEMS Research at Case Western Reserve University
MicroElectromagnectics Device Group at the University of Texas
MOSIS Fabrication Service of USC/Information Sciences Institute
Microsensos & Actuator Technology (MAT), Berlin
MiSTIC- Michigan Synthesis Tools for IC's
Nanofabrication Research Facility, UCLA Los Angeles
NanoNet, National Nanofabrication User Network
Ohio MEMSNet
RACE, The University of Tokyo
The Stanford Center for Design Research (CDR)
University of Sheffield, UK, MEMS Unit Homepage

Other addresses/institutions
Forschungszentrum Karlsruhe, Micro Systems Program (PMT)
http://pmtsun3.fzk.de/English/home_e.html

Forschungszentrum Karlsruhe, Institute for Microstructure Technology (IMT)
http://hbksun17.fzk.de:8080/IMT/eimt.htm
Sandia National Laboratories

http://www.mdl.sandia.gov/Micromachine/

Berkeley Sensor & Actuator Center
http://bsac.eecs.berkeley.edu/

University of Wisconsin at Madison
http://mems.engr.wisc.edu/

University of California at Los Angeles
http://synergy.icsl.ucla.edu/

MCNC MEMS Technology Applications Center
http://mems.mcnc.org/

Institute of Microtechnology GmbH, Mainz
http://www.uni-mainz.de/IMM/

The Microsystems Technology Laboratories (MTL) at MIT
http://goesser.mit.edu/

University of Bonn, Synchrotron Radiation Group, Prof. J. Hormes
http://syli34.physik.uni-bonn.de/rtlplus.html

MICROMACHINE CENTER
http://www.iijnet.or.jp/MMC/

The LIGA Club
http://www.dl.ac.uk/LIGA/

U.S. Synchrotron Sources
CAMD (Louisiana State University, Baton Rouge,LA)
ALS (LBL, Berkeley, CA USA)
APS (Chicago, IL USA)
NSLS I(BNL, Brookhaven, NY USA)
NSLS II(BNL, Brookhaven, NY USA)
SSRL (Stanford, USA)
UW Synchrotron Radiation Center (Stoughton, WI USA)

References

1. E. W. Becker, H. Betz, W. Ehrfeld, W. Glashauser, A. Heuberger, H. J. Michel, D. Münchmeyer, S. Pongratz, and R. V. Siemens, *Naturwissenschaften* **69**, 520 (1982).

2. P. Bley and J. Mohr, "The LIGA Process—a microfabrication technology," Research and Development Association for Future Electron Devices (FED), *FED Journal*, **5**, Suppl.1, 34–48 (1994).
3. W. Ehrfeld, P. Bley, F. Götz, P. Hagmann, A. Maner, J. Mohr, H. O. Moser, D. Münchmeyer, W. Schelb, D. Schmidt, and E. W. Becker, in *Proc. IEEE*, K. J. Gabriel and W. S. N. Trimmer, eds., Institute of Electrical and Electronic Engineers, 87TH0204-8 (1987).
4. R. Schenk, O. Halle, K. Müllen, W. Ehrfeld, and M. Schmidt, "Highly sensitive resist material for deep x-ray lithography,"*Proc. Micro- and Nano-Engineering '96* (1996).
5. Product information JENOPTIK Mikrotechnik GmbH, D-07739 Jena.
6. A. Schmidt, "Röntgentiefenlithographische Mikrostrukturfertigung Elektroneninduzierte Sekundäreffekte und Mehrfachbelichtungsprozeß," Ph.D. thesis, University of Bonn (1996).
7. G. Feiertag, *Röntgentiefenlithographische Mikrostrukturfertigung: Genauigkeit der Abbildung und Strukturierung durch Schrägbelichtung*, VDI report No. 242, Düsseldorf (1996).
8. Y. Vladimirsky, K. Morris, J. M. Klopf, O. Vladimirsky, and V. Saile, "X-ray micro lithography exposure system for high aspect ratio micromachining," *Proc. SPIE* **2650**, (1995).
9. B. Lai, D. C. Mancini, W. Yun, and E. Gluskin, "Beamline and exposure station for deep x-ray lithography at the advanced photon source,"*Proc. SPIE* **2880**, 171–176 (1996).
10. C. Khan Malek, R. Wood, B. Dudley, Z. Ling, and V. Saile, "Multi-exposure capability development for deep x-ray lithography for MEMS," *Proc. SPIE* **2880**, 160–170 (1996).
11. P. Bley and W. Menz, "High aspect ratio microstructure technology," Abstr. *HARMST '95*, Karlsruhe (1995).
12. Y. Vladimirsky, O. Vladimirsky, V. Saile, K. Morris, and J. M. Klopf, "Transfer mask for high aspect ratio micro-lithography," *Proc. SPIE* **2437**, (1995).
13. S. Ballandras et al., "Deep etch x-ray lithography using silicon gold masks fabricated by deep etch uv lithography and electroforming," *J. Micromech. Microeng.* **5**, 203–208 (1995).
14. H. Hein, P. Bley, J. Göttert, and U. Klein, S. 75, VDI Berichte 960, VDI-Verlag, Düsseldorf, (1992).
15. K. Kadel, Diplomarbeit, Universität Karlsruhe, unveröffentlicht (1987).
16. P. Coane and C. Friedrich, "Fabrication of composite x-ray masks by micromilling," *Proc. SPIE* **2880**, 130–141 (1996).
17. K. Kadel, G. Stern, R. Broßardt, and W. Schomburg, *Entwicklung eines subtraktiven Strukturierungsverfahrens zur Herstellung von Röntgenzwischenmasken für das LIGA-Verfahren*, KfK report 5262, Karlsruhe (1993).

18. H. Guckel, "Deep x-ray lithography for micromechanics and precision engineering," in *SRI '95*, (1995).
19. W. K. Schomburg, H. J. Baving, and P. Bley, "Ti and Be x-ray masks with alignment windows for the LIGA process," *Microelectron. Eng.* **13**, 323–326 (1991).
20. N. Gopinathan, "Direct fabrication of x-ray masks for LIGA by micromilling," Master's thesis, Louisiana Tech University, College of Engineering (1997).
21. H. Guckel, "LIGA and LIGA-like processing with high energy photons," *Microsystems Technology,* **(2)3**, 153–156 (1996).
22. J. Mohr, W. Ehrfeld, and D. Münchmeyer, KfK bericht 4414, Kernforschungszentrum, Karlsruhe (1988).
23. Formation of Microstructures Using a Preformed Photoresist Sheet, U.S. Patent #5378583 (1995).
24. A. Müller, J. Göttert, J. Mohr, and A. Rogner, "Fabrication of stepped micro optical benches for fibre and free-space applications," *Microsystems Technol.* **(2)** 40–45 (1996).
25. O. Schmalz, M. Hess, and R. Kosfeld, "Structural changes in poly(methyl methacrylate) during deep-etch x-ray synchrotron radiation lithography," *Die Angewandte Makromolekulare Chem.* **239**, Parts 1–3 (1996).
26. W. Ehrfeld, H. J. Baving, D. Beets, P. Bley, F. Götz, J. Mohr, D. Münchmeyer, and W. Schelb, *J. Vac. Sci. Technol.* **B6** (1), 178 (1988).
27. M. Born and E. Wolf, *Principles of Optics*, Pergammon Press, Oxford (1980).
28. E. E. Koch, D. E. Eastman, and Y. Farge, "Synchrotron radiation—a powerful tool in science," *Handbook on Synchrotron Radiation*, E. E. Koch, Ed., Vol. 1a, Chap. 1, Oxford University Press (1983).
29. S. Krinsky, M. L. Perlman, and R. E. Watson, "Characteristics of synchrotron radiation and of its sources," in *Handbook on Synchrotron Radiation*, E.E. Koch, Ed., Vol. 1a, Chap. 2, Oxford University Press (1983).
30. J. Schulz, A. El-Kholi, J. Göttert, K. Kadel, and J. Mohr, "The influence of sloped absorber sidewalls in deep x-ray lithography," *Microelectron. Eng.* **21** (1993).
31. G. Feiertag, M. Schmidt, and A. Schmidt, "Thermoelastic deformations of masks for deep x-ray lithography," *Microelectron. Eng.* **27** (1995).
32. T. Matsushita and H, Hashizume, "X-ray monochromators," in *Handbook on Synchrotron Radiation*, Vol. 1, North-Holland Publishing, 265–267 (1983).
33. R. Ghodssi, V. White, C. Herdey, C. Archie, D. D. Denton, and L. McCaughan, "An optimized LIGA/MEMS process for polymer waveguides on a linear microactuator," *Electrochem. Soc. Proc.* **95**, (1995).
34. A. Both, W. Bacher, M. Heckele, K. D. Müller, R. Ruprecht, and M. Strohrman, "Molding process with high alignment precision for the LIGA technology," in *Proc. MEMS* (1995).
35. A. Müller, J. Göttert, and J. Mohr, *Aufbau hybrider mikrooptischer*

Funktionsmodule für die optische Nachrichtentechnik mit dem LIGA-Verfahren, FZKA report 5786, Karlsruhe (1996).

36. Y. Vladimirsky, O. Vladimirsky, K. Morris, and G. Calderon, "X-ray microfabrication of multilevel structures and 3D patterning," in *Proc. MNE '96* 101–102 (1996).
37. J. Götter and J. Mohr, *Grundlagen und Anwendungsmöglichkeiten der LIGA-Technik in der Mikrooptik,* KfK report 5153, Karlsruhe (1993).
38. J. Göttert, M. Fischer, and A. Müller, "high-aperture surface relief microlenses fabricated by x-ray lithography and melting EOS," in *Topical Meeting Dig. Series* Vol. 5, *Microlens Arrays*, Teddington (1995).
39. F. J. Pantenburg and J. Mohr, "Influence of secondary effects on the structure quality in deep x-ray lithography," *Nuclear Instrum.Meth. Phys. Res. B* **97** (1995).
40. C. Müller and J. Mohr, *Miniaturisiertes Spektrometersystem in LIGA-Technik,* FZKA report 5609, Karlsruhe (1995).
41. H. Kalb, W. Bacher, M. Harmening, and R. Ruprecht, *Grundlegende Untersuchungen zur Fertigung gestufter Mikrostrukturen aus fluorierten Polymeren nach dem LIGA-Verfahren,* FZKA report 5576, Karlsruhe (1995).
42. H. Lehr, M. Schmidt, and G. Feiertag, "LIGA—only limited by our imagination," *LIGA News* **4**, April (1996).
43. J. Jara-Almonte, C. Friedrich, and R. Warrington, "Micromanufacturing," in *Handbook of Design, Manufacturing, and Automation*, R. Dorf and A. Kusiak Eds., Chap. 20, Wiley, New York (1994).
44. C. R. Friedrich and M. J. Vasile, "Precision Micromachining," in *The Industrial Electronics Handbook*, CRC Press LLC, 1505–1514 (1997).
45. M. Vasile, C. Friedrich, B. Kikkeri, and R. McElhannon, "micron-scale machining: tool fabrication and initial results," *Precision Eng.* **(119)** 2/3, 180–186 (1996).
46. M. Shaw, *Metal Cutting Principles*, Oxford Univ. Press, Oxford (1984).
47. G. Boothroyd and W. Knight, *Fundamentals of Machining and Machine Tools*, Marcel Dekker, New York (1989).
48. D. Lucca, R. Rohrer, and R. Komanduri, "Energy dissipation in the ultraprecision machining of copper," *Ann. CIRP* **(40)**1, 69–72 (1991).
49. E. Paul, C. Evans, A. Mangamelli, M. McGlauflin, and R. Polvani, "Chemical aspects of tool wear in single point diamond turning," *Precision Eng.* **(18)**1, 4–19 (1996).
50. C. Evans, "Cryogenic diamond turning of stainless steel," *Ann. CIRP* **(40)**, 571–575 (1991).
51. W. Blackley and R. Scattergood, "Ductile-regime machining model for diamond turning of brittle materials," *Precision Eng.* **(13)**2, 95–103 (1991).
52. T. Shibata, S. Fujii, E. Makino, and M. Ikeda, "Ductile-Regime Turning Mechanism of Single-Crystal Silicon," *Precision Eng.* **(18)**2/3, 129–137 (1996).

53. W. Bier, W. Keller, G. Linder, K. Schubert, and D. Seidel, "Manufacturing and testing of compact heat exchangers with high volumetric heat transfer coefficients," in *Microstructures, Sensors, and Actuators,* Vol.19, American Society of Mechanical Engineers DSC (1990).
54. W. DeVries, *Analysis of Material Removal Processes*, Springer-Verlag, New York (1992).
55. E. DeGarmo, J. Black, and R. Kosher, *Materials and Processes in Manufacturing*, Macmillan, New York (1984).
56. C. Friedrich and R. Warrington, "Techniques for micromechanical machining," in *Proc. Tooling '97*, 1997 (in press).
57. T. Masuzawa and M. Fujino, "A process for manufacturing very fine pin tools," Society of Manufacturing Engineers, Technical Paper MS90-307 (1990).
58. T. Masaki, K. Kawata, and T. Masuzawa, "Micro electro-discharge machining and its applications," *Proc. IEEE MEMS*, 21–26 (1990).
59. C. Friedrich and M. Vasile, "Development of the micromilling process for high-aspect-ratio microstructures," *J. MEMS* **(5)**1, 33–38 (1996).
60. P. Coane and C. Friedrich, "Fabrication of composite x-ray masks by micromilling," *Proc. SPIE* **2880**, 130–141 (1996).
61. C. R. Friedrich, P. J. Coane, J. Goettert, and N. Gopinathin, "Metrology and quantification of micromilled x-ray masks and exposures," *Proc. SPIE* **3048** (1997).
62. C. Friedrich and M. Vasile, "The micromilling process for high aspect ratio microstructures," *Microsystem Technol.* **(2)**3, 144–148 (1996).
63. E.W. Becker, W. Ehrfeld, P. Hagmann, A. Maner, and D. Münchmeyer, "Fabrication of microstructures with high aspect ratios and great structural heights by synchrotron radiation lithography, galvanoforming, and plastic molding (LIGA process)," *Microelectron. Eng.* **4**, 35 (1986).
64. P. Bley and J. Mohr, "The LIGA process—a microfabrication technology," *FED J.* **5**, Suppl. 1, 34 (1994).
65. W. Bacher, W. Menz, and J. Mohr, "The LIGA technique and its potential for microsystems," *IEEE Trans. Ind. Electr.* **42**(5), 431 (1995).
66. W. Ehrfeld, "Mikrosystemtechnik in Deutschland—Schaffen wir den Return on Research?" GIT 2/97 116 (1997).
67. O. Larsson, "Polymeric Microstructures and Replication Techniques," in *Proc. Microstructure Workshop*, Chap. 5.1 (1996).
68. *Statusbericht des Projekts Mikrosystemtechnik im Forschungszentrum Karlsruhe*, FZKA-bericht 5670 (1995).
69. A. Rogner, J. Eicher, D. Münchmeyer, R. P. Peters, and J. Mohr, "The LIGA technique—what are the new opportunities?" *J. Micromech. Microeng.* **2** 133 (1992).
70. R. Wechsung, "Mikrostrukturprodukte aus Metallen, Kunststoffen und Keramik," *Sensor Report* **1** 41 (1993).

71. R. Ruprecht, W. Bacher, J. H. Haußelt and V. Piotter, “Injection molding of LIGA and LIGA similar microstructures using filled and unfilled thermoplastics,” *Proc. SPIE* **2639** 146 (1995).
72. L. Weber, W. Ehrfeld, H. Freimuth, M. Lacher, H. Lehr, and B. Pech, “Micro molding—powerful tool for the large scale production of precise microstructures,” in *Proc. Symp. Micromachining Microfabrication,* 156 (1996).
73. R. Ruprecht, K. Müller, J. Konys, V. Piotter, A. Thies, H. Wöllmer, and J. Haußelt, “Various replication techniques for manufacturing three-dimensional metal microstructures,” *HARMST* (in press).
74. W. Bacher, K. Bade, K. Leyendecker, W. Menz, W. Stark, and A. Thommes, “Electrodeposition of microstructures: an important process in microsystem technology,” in *Electrochemical Technology*, N. Masuko, T. Osaka, and Y. Ito, Eds., p. 159, Kodansha/Gordon&Breach, Tokyo (1996).
75. F. Nöker and L. Keydel, “Herstellen von Mikrostrukturkörpern aus Kunststoffen,” *Kunststoffe* **82** 798 (1992).
76. J. Göttert, C. Müller and J. Mohr, “Examples and potential applications of LIGA components in microoptics,” in *Teubner Taschenbücher der Physik*, W. Ehrfeld et al., eds., Teubner, Stuttgart (1992).
77. W. Michaeli and A. Rogalla, “Spritzgießen von Mikrostrukturen,” *F&M Feinwerktechnik* **104**, 9, 641 (1996).
78. V. Piotter, T. Hanemann, R. Ruprecht, and J. Haußelt, “Injection molding and related techniques for fabrication of microstructures,” *Microsystem Technol.* **64** (1997) (to be published).
79. F. Nöker, private communication (1997).
80. A. Thommes, W. Bacher, K. Leyendecker, W. Stark, H. Liebscher, and Ch. Jakob, “LIGA microstructures from a NiFe alloy: preparation by electroforming and their magnetic properties,” in *Proc. 185th Meeting of the American Electrochem. Soc.*, 94–96 (1994).
81. M. Harmening, W. Bacher, P. Bley, A. El-Kholi, H. Kalb, B. Kowanz, W. Menz, A. Michel, and J. Mohr, “Molding of three-dimensional microstructures by the LIGA process,” in *Proc. MEMS '92*, p. 202, IEEE, 0-7803-0497-7/92 (1992).
82. K. Schubert, W. Bier, G. Linder, and D. Seidel, “Profiled diamonds for producing microstructures,” *Ind. Diamanten Rundschau* **24,** 235 (1990).
83. L. Lin, Ch-J. Chiu, W. Bacher, and M. Heckele, “Microfabrication using silicon mold inserts and hot embossing,” in *Proc. 7th Intern. Symp. Micro Machine and Human Science*, 66 (1996).
84. R. Ruprecht, T. Hanemann, and V. Piotter, “Polymer materials for microsystem technology,” in *Micro Materials '97* (1997) (to be published).
85. M. Heckele and W. Bacher, “Modular molding equipment for quasi monolithic integration of micromechanical and microelectronic components,” in *Mechatronics*, 501 (1996).

86. M. Heckele and W. Bacher, "Modular Hot embossing equipment for MEMS," *Micromachine Devices* **2**(2), 1, (1997).
87. T. Hanemann, V. Piotter, R. Ruprecht, and J. Haußelt, "Molding of polymeric microstructures," in *Bayreuth Polymer Symposium*, (1997).
88. B. Büstgens, W. Bacher, W. Bier, R. Ehnes, D. Maas, R. Ruprecht, and W.K. Schomburg, "Micromembrane pump manufactured by molding," in *Proc. Actuator '94*, 86 (1994).
89. J. Fahrenberg, W. Bier, D. Maas, W. Menz, R. Ruprecht, and W.K. Schomburg, "Microvalve system fabricated by thermoplastic molding," in *Proc. Micro Mechanics Europe,* 17 (1994).
90. C. Goll, W. Bacher, B. Büstgens, D. Maas, R. Ruprecht, and W. K. Schomburg, "Electrostatically actuated polymer microvalve equipped with a movable membrane electrode," *Proc. MME '96*, 231 (1996).
91. J. Mohr, J. Göttert, and A. Müller, "Microoptical devices based on free space optics," *Proc. SPIE* **2783**, 6 (1996).
92. H.-D. Bauer, W. Ehrfeld, M. Gerner, T. Paatzsch, A. Picard, H. Schift, and L. Weber, "Microoptical elements and optical interconnection components fabricated using the LIGA technique," in *Proc. Int. Symp. on Microsystems, Intelligent Materials and Robots 1995*, J. Tani, Ed., 33 (1995).
93. M. T. Gale, L. G. Baraldi, and R. E. Kunz, "Replicated microstructures for integrated optics," *Proc. SPIE* **2213,** 2 (1994).
94. B. Kluepfel and F. Ross, (Eds.), *Holographic Marketplace*, p. 63, Ross Books, Berkeley, CA (1991).
95. R. Klein and A. Neyer, "Silicon micromachining for micro-replication technologies," *Electron. Lett.* **30** (20) 1672 (1994).
96. A. Müller, J. Göttert, J. Mohr, and A. Rogner, "Fabrication of stepped microoptical benches for fibre and free space application," *Microsystem Technol.* **2** 40–45 (1996).
97. O. Larsson, O. Ohman, A. Billman, L. Lundbladh, C. Lindell and G. Palmskog, "Silicon based replication technology of 3D-microstructures by conventional CD-injection molding techniques," *Proc. Transducers '97, International Conference on Solid-State Sensors and Actuators*, 1415–1418 (1997).
98. R. Wiesmann, S. Kalveram, S. Rudolph, M. Johnck, and A. Neyer, "Single-mode polymer waveguides for optical backplanes," *Electron. Lett.* **32** (25) 2329–2330 (1996).
99. J.-H. Kim and A. Neyer, "Design and simulation of polymeric waveguide devices fabricated by replication technologies," *J. Opt. Commun.* **17** 172–178 (1996).
100. U. Köhler, A. E. Guber, W. Bier, M. Heckele, and Th. Schaller, "Fabrication of microlenses by combining silicon technology, mechanical micromachining and plastic molding," *Proc. SPIE* **2687** 18 (1996).
101. C. Ossmann, J. Goettert, M. Ilie, J. Mohr, and P. Ruther, "Fabrication of

PMMA based microlenses using the LIGA process," *Proceedings EOS Topical Meeting on Microlens Arrays*, (1997).
102. Fischer, J. Goettert, A. E. Guber, U. Koehler, J. Mohr, F.-W. Oertmann, and P. Ruther, "Auslegung und Herstellung von Kunststoff-Mikrolinsen und Mikrolinsensystemen fur flexible Endoskopie in der Neurochirugie," *VDI-Verlag, Series Innovationen in der Mikrosystemtechnik,* **50,** 107–125 (1997).
103. C. Müller and J. Mohr, "Microspectrometer fabricated by the LIGA process," *Interdisciplinary Sci. Rev.* **18** (3), 273 (1993).
104. P. Krippner, J. Mohr, C. Müller, and C. van der Seel, "Microspectrometer for the infrared range," in *Proc. Microoptical Technologies for Measurement, Sensors, and Microsystems*, 277 (1996).
105. A. Both, W. Bacher, and R. Ruprecht, "Quasimonolithic integration of LIGA microstructures with integrated circuits," *Proc. MME '93*, 93 (1993).
106. M. Strohrmann, P. Bley, O. Fromhein, and J. Mohr, "Acceleration sensor with integrated compensation of temperature effects fabricated by the LIGA process, *Sensors Actuators A* **41–42** 426 (1994).
107. B. Rogge, J. Schulz, J. Mohr, A. Thommes, and W. Menz, "Magnetic microactuators fabricated by the LIGA-Technique for Large Displacements or Large Forces," in *Proc. Actuator '96*, 112 (1996).
108. E. Günther and U. Maciejewski, 8th CIMTEC, Florence, 28.6. - 4.7. (1994).
109. R. Knitter and C. Odemer, *KfK-Nachrichten* 240–245 (1994).
110. R. Knitter, E. Günther, U. Maciejewski, and C. Odemer, cfi/Ber. DKG **71**, No. 9, 549–556 (1994).
111. W. Bier, G. Linder, and D. Seidel, *KfK-Nachrichten* **23** 165–173 (1991).
112. P. Bley and W. Menz, *KfK-Nachrichten* **23**, 69–75 (1991).
113. H.-J. Ritzhaupt-Kleissl, W. Bauer, E. Günther, J. Laubersheimer, and J. Hauelt, *HARMST 95*, (in press).
114. W. Bauer, H.-J. Ritzhaupt-Kleissl, and J. Hauelt, Werkstoffwoche, Stuttgart (1996).
115. J. Laubersheimer, S. Röbel, and H.-J. Ritzhaupt-Kleissl, Werkstoffwoche, Stuttgart (1996).
116. R. Knitter, E. Günther, C. Odemer, and U. Maciejewski, "Ceramic microstructures and potential applications," *Microsystem Technologies* **(2)**3 135–138 (1996).
117. W. A. Smith., A. A. Shaulov, and B.A. Auld, *Ferroelectrics* **91** 155–162 (1989).
118. R. Knitter and E. Günther, Werkstoffwoche, Stuttgart (1996).
119. E.W. Becker, W. Ehrfeld, and D. Munchmeyer, *Untersuchungen zur Abbildungsgenauigkeit der Runtgentiefenltihographie mit Synchrotronstrahlung bei der Herstellung technischer Trenndusenelemente*, KfK report 3732, Forschungszentrum Kalrsuhe (1984).

120. W.Schomburg, R. Ahrens, W. Bacher, S. Engemann, P. Krehl and J. Martin, "Long-term performance analysis of thermo-pneumatic micropump actuators," *Proc. Transducers '97, International Conference on Solid-State Sensors and Actuators*, 365–368 (1997).

CHAPTER 7

Sensors and Actuators

Hal Jerman
Quinta Corporation

Stephen Terry
EG&G IC Sensors

CONTENTS

7.1 Introduction / 381

7.2 Sensor Classification / 381

7.3 Mechanical Properties of Silicon and Thin Film Materials / 382
  7.3.1 Silicon elastic modulus and expansion coefficient / 383
    7.3.1.1 Thermal expansion coefficient / 383
  7.3.2 The piezoresistive effect in silicon / 384

7.4 Pressure Sensors / 385
  7.4.1 Piezoresistive pressure sensors / 386
  7.4.2 Capacitive pressure sensors / 389
  7.4.3 Resonant pressure sensors / 390
  7.4.4 Future trends / 391

7.5 Accelerometers / 392
  7.5.1 Capacitive accelerometers / 394
    7.5.1.1 Surface-micromachined capacitive accelerometers / 396
  7.5.2 Force-balanced accelerometers / 397
  7.5.3 Resonant accelerometers / 399
  7.5.4 Tunneling accelerometers / 399
    7.5.4.1 Alternative tunneling structures / 400
  7.5.5 Conclusions / 401

7.6 Flow Sensors / 401
  7.6.1 Introduction / 401
  7.6.2 Gas flow sensors / 402
  7.6.3 Surface shear sensors / 405
  7.6.4 Humidity sensors / 405
  7.6.5 Liquid flow sensors / 405
    7.6.5.1 Thermal dilution flow sensor / 407
    7.6.5.2 Thermal transit-time sensor / 407
  7.6.6 Differential pressure flow sensors / 408
  7.6.7 Conclusions / 408

7.7 Actuators / 408
    7.7.1 Introduction / 408
    7.7.2 Microvalve examples / 410
        7.7.2.1 Microvalves with external actuators / 410
        7.7.2.2 Microvalves with integrated actuators / 413
    7.7.3 Check valves / 420
    7.7.4 Micropumps / 421
        7.7.4.1 Mechanical pumps / 421
        7.7.4.2 Nonmechanical pumps / 425
    7.7.5 Conclusions / 427

References / 427

7.1 INTRODUCTION

Microsensors have become an essential element in many measurement systems, finding countless applications in industrial monitoring, factory automation, the automotive industry, transportation, telecommunications, computers and robotics, environmental monitoring, health care, and biotechnology. A substantial driving force behind this progress comes from developments in semiconductor fabrication technology and in electronic signal processing. With the development of microprocessors and application-specific integrated circuits, signal processing has become cheap, accurate, and reliable. Sensors continue to lag behind electronics in terms of cost/performance ratios, however, due largely to the lack of general-purpose sensor elements that can be manufactured and packaged in the same way as electronic components.

There are many commercial opportunities in the sensor area, and many new sensor technologies and devices are being developed. New products and companies are emerging from this effort, stimulating further investment in this area. Application of sensors brings new dimensions to products in the form of performance, convenience, cost savings, and increased safety.

7.2 SENSOR CLASSIFICATION

Sensing the real world requires dealing with physical and chemical quantities that are diverse in nature. From the measurement point of view, physical and chemical quantities have been conventionally divided into six signal domains.[1]

1. The thermal signal domain, including temperature, heat, and heat flow.
2. The mechanical signal domain, such as force, pressure, velocity, acceleration, and position.
3. The chemical signal domain, measuring the internal quantities of matter such as concentration of a material, composition, or reaction rate.
4. The magnetic signal domain, including magnetic field intensity, flux density, and magnetization.
5. The radiant signal domain: the signals are quantities of the electromagnetic waves such as intensity, wavelength, polarization, and phase.
6. The electrical signal domain, measuring voltage, current, and charge.

Sensors convert nonelectrical physical or chemical quantities into electrical signals. The principle of operation of a particular sensor depends on the type of physical quantity it is designed to sense. The classification of sensors can follow the classification of physical quantities the device is designed to sense, or sensors can be classified based on their use of external energy sources. Sensors that generate an electrical output signal without an additional energy source are called passive, such as a thermocouple, while sensors that generate an electrical output signal with the help of an auxiliary energy source are called active. Sensors such as magnetotransistors and phototransistors operate by modulating an external energy source that serves as the main source for the output signal. There are a large number of well-known effects, such as the piezoelectric effect or the Seebeck effect, that can be used to convert one type of physical signal to another, and compilations of these effects are available for reference.[2]

7.3 MECHANICAL PROPERTIES OF Si AND THIN FILM MATERIALS

Fundamental to the design of micromachined devices are the interactions between the structure, layout, fabrication, and material properties of the device. Often the fundamental properties or limits of the sensor or actuator are due to the properties of the structural supports. Many of the advantages of micromachined devices are due to the exceptional mechanical characteristics of silicon and the thin film materials used as a part of the fabrication technology. Some of the properties of some of these materials are shown in Table 7.1.[3]

TABLE 7.1 Silicon and thin film mechanical properties

Material	Yield (GPa)	Modulus (GPa)	Density (g/cm^3)	Thermal Conductivity (W/cm°C)	Thermal Expansion (xE-6/°C)
Diamond	53	1035	3.5	20	1.0
SiC	21	700	3.2	3.5	3.3
Al_2O_3	15.4	530	4.0	0.5	5.4
Si_3N_4	14	385	3.1	0.19	0.8
SiO_2	8.4	73	2.5	0.014	0.55
Si	7.0	190	2.3	1.57	2.33
Steel	4.2	210	7.9	0.97	12
W	4.0	410	19.3	1.78	4.5
Mo	2.1	343	10.3	1.38	5.0
Al	0.17	70	2.7	2.36	22
Au	0.13	57	19.3	2.97	14.2

Silicon itself has elastic properties similar to steel, but with substantially lower density and higher thermal conductivity. One of its primary limitations, compared with traditional metallic mechanical materials, is the mechanical failure of silicon and thin film dielectrics by way of brittle fracture. Thus, while problems with mechanical hysteresis tend to be minimal in silicon-based devices, often elaborate mechanical protection mechanisms, such as limit stops and damping, must be included in the device structure to prevent the breakage of suspension elements due to shock.

7.3.1 Silicon Elastic Modulus and Expansion Coefficient

The design of high-performance silicon mechanical sensors, resonators, and spring systems requires the availability of accurate data on the elastic properties of silicon. The variations of these properties with temperature are important for the development of mechanical microstructures with well-defined thermal characteristics.

Monocrystalline silicon belongs to the m3m class of the cubic crystal system. This is in contrast to most conventional mechanical materials which are isotropic, where the mechanical properties of the material are independent of direction. Silicon is characterized by three independent elastic stiffness parameters: c_{11}, c_{44}, and c_{12}, corresponding to orthogonal directions in the crystal lattice. The various vibration modes of silicon resonators, as well as their thermal properties, generally depend on several elastic constants, and on their orientation relative to the crystallographic axes. Recently, the temperature coefficients of the three independent elastic constants of silicon have been measured by analyzing the frequency-temperature curves of four high-performance silicon resonators.[4]

7.3.1.1 Thermal expansion coefficient

For the calculation of the temperature coefficients of elasticity, precise values of the thermal expansion coefficients of silicon are necessary. The first-, second-, and third-order coefficients have been fitted to reported data[5] on thermal expansion measured at different temperatures. Based on these data values between -93 °C and +106 °C, the coefficients of the thermal expansion and the temperature coefficients of density at 25 °C have been determined and are summarized in Table 7.2.

TABLE 7.2 Calculated elastic stiffness $c_{ij.k}$, thermal expansion $a_{ij.k}$, and temperature coefficients of density $Tr^{(k)}$ at 25 °C for 4 Ω-cm p-type material.

c_{11}	165.6×10^{9} N/m^2
c_{44}	79.5×10^{9} N/m^2
c_{12}	63.9×10^{9} N/m^2
$c_{11\,1}$	-80.5×10^{-6}/°C
$c_{44.1}$	-58.2×10^{-6}/°C
$c_{12\,1}$	-115.0×10^{-6}/°C
$a_{11\,1}$	2.84×10^{-6}/°C
$a_{44\,1}$	8.5×10^{-9}/°C
$a_{12\,1}$	-32×10^{-12}/°C
$Tr^{(1)}$	-8.5×10^{-6}/°C
$Tr^{(2)}$	-25.2×10^{-9}/°C
$Tr^{(3)}$	95.3×10^{-12}/°C

7.3.2 The Piezoresistive Effect in Silicon

The piezoresistive coefficients in silicon have been investigated by many authors. Smith[6] discovered the piezoresistance effect of semiconductors having anisotropic energy band structures and published experimental data for lightly doped bulk silicon, and Tufte, Chapman, and Long[7] found that the piezoresistive coefficient decreases with increasing surface concentrations in diffused layers. The piezoresistance coefficients, usually required in an arbitrarily oriented coordinate system, can be calculated from a tensor transformation for the particular orientation.

The resistance R_0 of a rectangular conductor is expressed by[8]

$$R_0 = \frac{\rho_0 l}{wt} , \tag{7.1}$$

where ρ_0 is the resistivity and l, w, and t are the length, width, and thickness of the conductor, respectively. When the resistor is stretched in length, the relative change in resistance is given by

$$\frac{\Delta R}{R} = \frac{\Delta l}{l} - \frac{\Delta w}{w} - \frac{\Delta t}{t} + \frac{\Delta \rho}{\rho_0} . \tag{7.2}$$

Introducing Poisson's ratio ν, where

$$\frac{\Delta w}{w} = \frac{\Delta t}{t} = -\nu \frac{\Delta l}{l} , \tag{7.3}$$

the gauge factor G (strain sensitivity) is

$$G = \frac{\Delta R / R_0}{\varepsilon} = 1 + 2\nu + \frac{\Delta \rho / \rho_0}{\varepsilon} , \tag{7.4}$$

where $\epsilon = \Delta l / l$ is the strain. The first two terms represent the change in resistance due to dimensional changes (dominant for metal gauges), while the last term represents the change in resistivity (dominant for semiconductor gauges). In semiconductor gauges, the resistivity change is larger than the dimensional change by about a factor of 50, and the dimensional change is generally neglected. If a uniaxial stress T is applied along the direction of current flow, the piezoresistance ratio is given by

$$\frac{\Delta R}{R_0} = \frac{\Delta\rho}{\rho_0} = \pi_1 T \quad , \tag{7.5}$$

where π_1 is the longitudinal piezoresistance coefficient and T is the longitudinal stress. The relation between π_1 and G is

$$G = Y_1 \pi_1 \quad , \tag{7.6}$$

where Y_1 is Young's modulus and is anisotropic with respect to crystal direction.

Much of the theoretical understanding of the piezoresistance effect is based on the early work of Herring,[9] which describes the effect of lattice strain on the perturbation of the scattering of carriers from one energy valley to another, which changes the averaging over longitudinal and transverse effective masses. Kanada has published the results of calculations of the piezoresistive coefficients as functions of both temperature and surface concentration.[10] For p-type silicon diffused layers, π_{44} is about 100×10^{-12} cm^2/dyn for a surface concentration of 10^{19} cm^{-3}. This is in reasonable agreement with experimental results.

7.4 PRESSURE SENSORS

Transducers for pressure are among the oldest to be manufactured in silicon[11] and are devices for which the demand continues to be very high. Specific designs and mounting arrangements vary considerably, ranging from the small, sensitive, catheter-tip transducers used within the heart to the larger, rugged devices needed for industrial process control. However, in spite of the wide range of applications, two approaches have come to dominate development efforts in silicon—the piezoresistive and capacitive.

Piezoresistive structures have been the traditional basis for pressure transducers, offering small size, excellent linearity over a wide dynamic range, moderately high pressure sensitivity, and relative freedom from hysteresis and creep. Piezoresistive and capacitive pressure sensors are quite different devices in that the former measure material stresses while the latter measure average deflection. While deflection and stress are directly coupled for uniformly applied pressure, they may not be in other situations, due to temperature or nonlinear deflection.

Several parallel-plate capacitive structures[12,13] have been described, and this approach has come to be preferred in some areas. Capacitive transducers have been shown to offer higher pressure sensitivity and lower temperature sensitivity than their piezoresistive counterparts; however, they also require a larger die area, are more nonlinear, and require more sophisticated sensing circuitry. There are several papers in the literature that attempt to compare the theoretical limitations of the two types of pressure sensors, using fundamental considerations such as thermal noise, Brownian motion of fluid particles, etc.[14,15] At present, the devices are not approaching these theoretical limits and there are additional practical considerations, such as packaging stress, side-wall touching effects, and flow sensitivity, which are difficult to estimate and have not been included in these comparisons. Thus it is likely that piezoresistive transducers will continue to see wide use in some areas even as capacitive structures are adopted for others.

7.4.1 Piezoresistive Pressure Sensors

It has been shown[16] that the use of p-type resistors oriented in the (110) direction on a (100) silicon diaphragm offers the highest stress sensitivity and greatest compatibility with standard wafer processing. Such resistors are normally arranged in a Wheatstone bridge so that the temperature coefficients of the individual resistors cancel (to the degree the resistors are matched).

Figure 7.1 illustrates a typical version of a piezoresistive pressure sensor. The chip consists of a thin silicon diaphragm that has been selectively etched from a much thicker silicon wafer. The thin diaphragm region is supported by a thick silicon "frame," which serves to clamp the edges of the diaphragm. To increase the stiffness of the frame and isolate the diaphragm from stresses coming from the adhesive attaching the sensor to its package, a glass or silicon constraint layer is bonded to the back side of the silicon sensor chip. This bonding process is typically performed at the wafer level before the sensor chips are sawn apart.

When a differential pressure is applied across the square silicon diaphragm shown in Fig. 7.2, the diaphragm deflects, resulting in a symmetric stress distribution with maximum values at the center of each edge of the diaphragm, as shown in Fig. 7.3. Piezoresistors are diffused into the diaphragm in those regions to detect the stress that results from the deflection of the diaphragm. By using resistors oriented parallel as well as normal to the diaphragm edges, both increasing and decreasing resistance with pressure can be obtained. By connecting opposite-responding resistors in opposing legs of a Wheatstone bridge, full bridge sensitivity can be obtained.

The rectangular diaphragm pressure sensor is an approximation of an infinitely long diaphragm. In this case with pressure applied to the upper surface of the diaphragm, the resulting stress varies from tensile to compressive along a path from the long edges in toward the center of the diaphragm. The resistors are all placed perpendicular to the strain field, but the outer resistors are in tension and the inner ones are in compression. As with the square diaphragm, with the resistors in opposing legs of a Wheatstone bridge, full bridge sensitivity is achieved with this

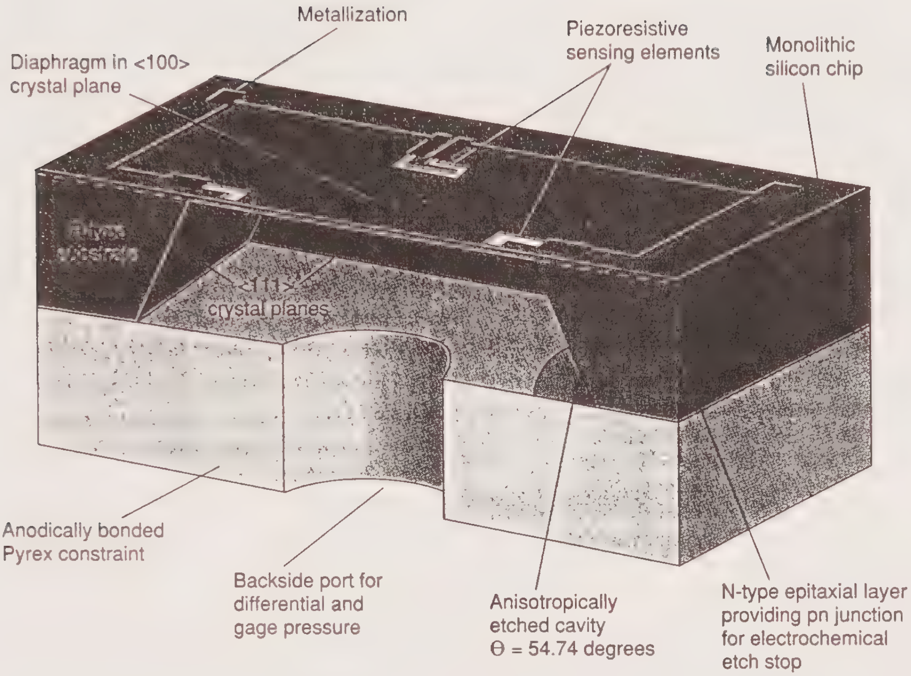

Metallization
Diaphragm in <100> crystal plane
Piezoresistive sensing elements
Monolithic silicon chip
P-type substrate
<111> crystal planes
Anodically bonded Pyrex constraint
Backside port for differential and gage pressure
Anisotropically etched cavity Θ = 54.74 degrees
N-type epitaxial layer providing pn junction for electrochemical etch stop

FIG. 7.1 A cross-sectional view of a typical bulk-micromachined pressure sensor. The thin silicon diaphragm is formed by etching the silicon substrate to leave a thin diaphragm. Diffused resistors sense the strain caused by deflection of the diaphragm. [103] (©Lucas NovaSensor)

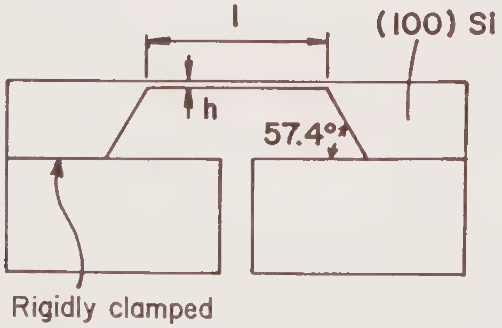

l
(100) SI
h
57.4°
Rigidly clamped

FIG. 7.2 Cross-section of a typical bulk-micromachined pressure sensor showing the side walls caused by anisotropic etching and the presumed clamped boundary conditions on the silicon sensor provided by a bonded constraint.[104] (©1986 IEEE)

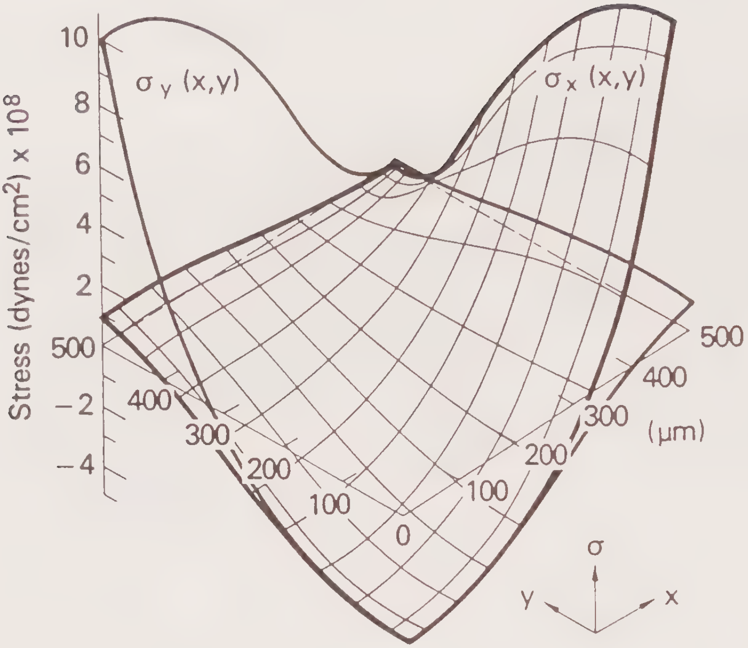


FIG. 7.3 Calculated stress distributions for the device shown in Fig. 7.2 with an applied pressure of 300 mmHg and diaphragm dimensions of 1 mm × 10 μm.[104] (©1986 IEEE)

configuration. Resistor placement in this type of sensor tends to be less critical than in a square diaphragm device.

If the cavity that is formed between the diaphragm and the constraint is sealed with a vacuum inside, the pressure sensor is an "absolute" pressure sensor, measuring the absolute pressure being applied to the outer surface of the diaphragm. If the cavity is vented to atmosphere, the sensor measures "gauge" pressure relative to atmospheric pressure. A cavity connected to another pressure source results in a differential pressure sensor.

The performance of the sensor varies over both temperature and pressure. The sensitivity of the piezoresistors decreases as temperature increases. Any residual stresses due to the bonding of the constraint layer or the die attach material which holds the chip to the package may also contribute to temperature-dependent changes in sensitivity and zero-pressure output from the sensor. The nonlinearity of response to applied pressure depends upon such factors as the location of the resistors in the strain field on the diaphragm and the deflection of the diaphragm. As the deflection increases towards 10% of the diaphragm thickness, the nonlinearity of the resulting strain increases as the diaphragm moves from the linear plate bending mode to nonlinear membrane deflection. Properly designed devices are capable of achieving nonlinearity less than 0.01% of full-scale output.

7.4.2 Capacitive Pressure Sensors

As shown in Fig. 7.4, a capacitive pressure transducer converts the diaphragm deformation, corresponding to pressure, into a change of capacitance, then into output electrical signals such as changes of oscillating frequency, time, charge, and voltage.[17,18] The total change of capacitance is the integral of the change of each small area on the diaphragm. The change of capacitance, $\Delta C/C$, is not linear with respect to deformation or pressure, but the relationship is reproducible and its full-scale value can be 20–200%. The structure is relatively simple and can be fabricated with conventional micromachining techniques. The disadvantage is that the capacitance is small (generally 1–3 pF). The measurement circuit has to be integrated on the chip or specially designed to null the stray capacitance; otherwise the parasitic capacitance associated with lead wires and bonding pads will generate significant interference and noise.

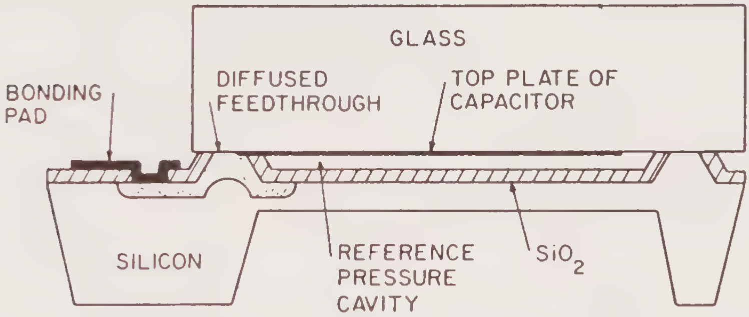


FIG. 7.4 Cross-section of a typical capacitive pressure sensor. The reference pressure is often a vacuum to reduce temperature effects caused by reference pressure expansion.[13] (©1982 IEEE)

A comparison between capacitive and piezoresistive transducers was made at Case Western Reserve University in 1980, where both types of pressure sensors were designed and tested. A capacitive device was found to have the potential for higher performance than piezoresistive devices in applications requiring high sensitivity, low pressure range, or very stable miniature sensors.

To achieve high sensitivity and wide operating range in a micromachined capacitive pressure sensor, one possible design approach is to use an ultrathin diaphragm with a center boss.[19] This approach has several advantages: (1) An ultrathin diaphragm provides very high sensitivity; (2) a center boss enlarges the operating range and yields improved linearity; and (3) a nearly parallel-plate capacitor is obtained using the center boss. Analysis has led to the following findings[20]: (1) The linearity of the load-deflection characteristic improves as the boss thickness increases. This improvement is minor when the boss thickness is less than five times the diaphragm thickness but becomes significant when the boss thickness is about ten times the diaphragm thickness because with that thickness ratio, the center boss can, like a plate, deflect uniformly under load. (2) There is a trade-off in selecting the best solidity ratio because it affects both the operating range and the pressure sensitivity. The center boss is designed to serve as a conducting plate for the capacitive pressure sensor; hence, when it is too small, a smaller than desired output signal results. On the other hand, too high a solidity ratio can make the diaphragm too stiff, degrading the sensitivity.

In one design, the device has a gap of 2.76 μm. The diaphragm is circular, 2 mm in diameter, and 0.3 μm in thickness. The center boss is also circular with a 1.2-mm-diameter and a thickness of 3 μm. This ultrasensitive capacitive pressure sensor design has been extensively tested. Figure 7.5 shows the measured capacitance change versus applied pressure. As can be seen, the capacitance change is highly linear because of the design of the center boss. Since the boss is ten times as thick as the diaphragm, most of the diaphragm deformation comes from stretching and bending of the dielectric layer. The minimum measured pressure sensitivity of the device is 5 fF/mtorr, and the equivalent boss deflection is 25 Å/mtorr. The minimum resolvable pressure is 0.1 mtorr. The operating range of the device has been studied from 0 to 1000 mtorr (0–130 Pa), which covers the range of sound pressure levels and complements existing barometric devices.

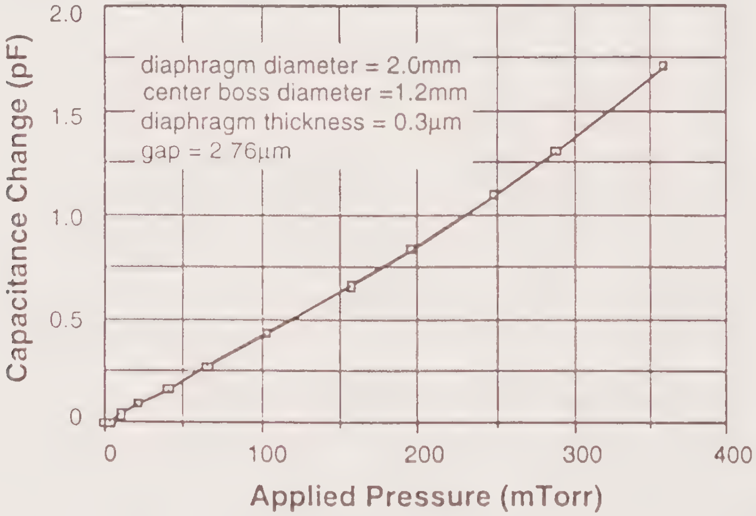


FIG. 7.5 Measured capacitance change with applied pressure for a bossed, circular diaphragm pressure sensor.[20] (©1994 IEEE)

7.4.3 Resonant Pressure Sensors

Improvements in precision silicon micromachining technology during the past decade have led to a wide range of improved sensing structures, including the use of resonant microbeams as strain transducers.[21,22] With resonant sensors, the output is in the form of a robust sinusoidal signal whose frequency changes with pressure. The output is quasi-digital in the sense that it can easily be converted to digital form using a high-frequency clock. Strain-sensitive mechanical-grade polysilicon resonators have been fabricated monolithically on silicon pressure diaphragms with integral vacuum-sealed cavities (see Fig. 7.6) for differential or absolute pressure sensors that are usable in a wide range of applications.[23] Simplified analysis of a resonant microbeam pressure sensor requires a model of the deflection profile of the pressure diaphragm under load, the resulting strain applied to the microbeam(s) as a function of the position on the diaphragm, and the shift in resonant frequency of the microbeam with applied strain. With these basic elements, the microbeam frequency response as a function of applied pressure can be predicted as a function of geometry and material properties.

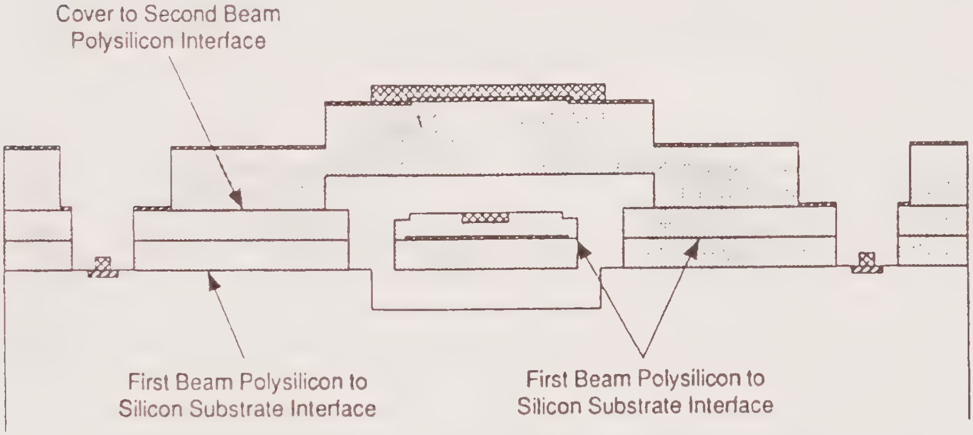

Cover to Second Beam
Polysilicon Interface
First Beam Polysilicon to
Silicon Substrate Interface
First Beam Polysilicon to
Silicon Substrate Interface

FIG. 7.6 Cross-section of a resonant microbeam sensor element showing the interfaces between the polysilicon layers. The resonant element is sealed in a vacuum cavity to maximize the Q of the mechanical resonance.[23] (©1992 IEEE)

7.4.4 Future Trends

The trends in the pressure-sensing industry for micromachined devices include increased integration; the development of higher operating temperature sensors; extended operating range of existing sensors; improved accuracy and resolution; more complex structures utilizing micromachining technology; cost reduction through process improvements and die size reduction; and improved packaging for smaller size, surface-mounted assemblies and improved media compatibility.

7.5 ACCELEROMETERS

Micromachined accelerometers represent a rapidly growing field both for commercial and technological development. All current accelerometers consist of a proof, or seismic mass supported by a suspension to a frame. Under the influence of an acceleration, the proof mass generates a force that results in a displacement or strain in the suspension element.

The means by which this force is measured are varied, depending on issues of performance and cost. Although the batch fabrication processes that revolutionized microelectronics hold out the promise of low-cost silicon accelerometers, fulfilling that promise while achieving reliable devices, even in the low-to-moderate performance range, has proved challenging. Among the challenges are optimization of the mechanical damping, temperature compensation, over-force stops for the proof mass to prevent damage due to shock, and decoupling the device from package-induced stresses. Overcoming these obstacles has required sophisticated mechanical modeling and, in some cases, development of new fabrication techniques.

A relatively new commercial field for micromachined accelerometers is in electronic systems for airbags. In the case of a central airbag system, the electronics are placed in the middle of the car rather than having individual sensors at the forward corners of the car. In the event of a crash, the acceleration sensor supplies an electrical signal that corresponds to the deceleration. Algorithms in the controller decide whether to fire the airbag. Accelerometers used in central airbag systems have been constructed by using a variety of different micromachining technologies. They have to be geometrically small and must be available in high production volumes. Although the quality requirements are very high, costs must be extremely low. Current pricing for a packaged device with amplified electrical output is less than $4 and the projected price is below $2 in the year 2000. The present market offers several acceleration sensors meeting the requirements for use in airbag systems.

Silicon microaccelerometers date from the 1970s. Early research focused on using diffused piezoresistors and anisotropic etching technologies that were developed for silicon diaphragm pressure sensors. The first silicon micromachined accelerometer used bulk micromachining with piezoresistors to sense acceleration.[24] The silicon mass was suspended by a single cantilever that determined the cross-axis sensitivity. Other bulk micromachined sensors contain a seismic mass suspended by one, two, or four silicon beams.[25] The beams carry piezoresistors, which in order to achieve a maximum output signal, are placed at the edges of the beams where the maximum bending strain occurs.

Later designs improved on the single cantilever configuration by using multiple suspensions to reduce cross-axis sensitivity, separate the various vibrational modes, and improve the robustness of the units. In an IC Sensors design,[26] two piezoresistors are diffused onto each of the four suspension cantilevers. These eight piezoresistors are connected with two resistors on each arm of a Wheatstone bridge so that the multimode vibration of the proof mass will be suppressed. To control damping, narrow air gaps surrounding the seismic mass in the direction of deflection are provided, or the sensors can be placed in oil-filled housings.[27]

Capacitive sensing of the proof mass displacement is attractive because it avoids the temperature dependence that is characteristic of piezoresistors. Several prototypes have been reported, with most attempting to achieve high performance through optimized suspensions and highly sensitive interface electronics, but at the cost of more elaborate fabrication processes and interface circuitry.

Accelerometers designed with a feedback actuator that can provide a restoring force to the proof mass can operate as a closed-loop sensor with a proof mass that moves negligibly in response to inertial forces. In this type of accelerometer, the electrical signal applied to the actuator is usually used as the sensor output signal. One early design used a separate torquer coil mounted to the proof mass to provide a magnetic restoring force and piezoresistive strain gauges to sense its position. A more convenient approach, for the case of capacitive position sensing, is to apply a signal to capacitor plates, which generates an electrostatic restoring force.[28] For capacitive-based accelerometers, the same structure can often be used in both open- and closed-loop modes, with the latter requiring additional signal processing circuitry to close the loop. There has recently been significant interest in this class of accelerometers for both consumer products and high-precision applications.[29,30]

An alternative approach to achieving high sensitivity and large dynamic range is to use a resonant microstructure as a strain sensing element. The frequency of a resonant structure can be made highly sensitive to compressive or tensile strain with careful design, and that resonant frequency can be measured with high precision. Resonant accelerometers are commercially available that use chemically milled quartz tuning forks as the resonant strain gauges. If the entire accelerometer consisting of proof mass, suspension, and micromechanical resonators can be fabricated in silicon, then substantial cost savings can be achieved.[31]

Two rather different micromachining technologies are capable of fabricating the proof mass, suspension, and limit stops to prevent breakage from shock. Bulk anisotropic etching of silicon is a well-developed technology and has long been used for silicon diaphragm pressure sensors. Accelerometers require more sophisticated anisotropic etching, such as double-sided etching, compensation of outside corners, or the use of p+ or electrochemical etch stops.

More recently, the fabrication of microstructures from deposited thin films, known as "surface micromachining," has developed into a competitive technology for accelerometer fabrication. It is well suited to the fabrication of force-balance accelerometers that use capacitive sensing and electrostatic feedback.[32] Because the proof mass is a suspended thin film plate, it occupies much less area than that required for the bulk micromachined silicon proof mass; this implies a significant cost advantage. Also for this reason there is much more incentive to integrate the sensing and feedback electronics with the suspended proof mass, a technique that may lead to further reductions in the cost of the accelerometer system.

7.5.1 Capacitive Accelerometers

In recent years, much effort has been focused on developing silicon-based capacitive accelerometers. In these devices, one of the two electrodes is the suspended proof mass, which deflects with respect to the fixed electrode when it accelerates. This capacitor is used in a capacitive bridge to transduce the displacement into a voltage. The fabrication of capacitive accelerometers can be very similar to that of piezoresistive accelerometers. A silicon wafer is etched from the back side (or both sides) to form the suspension and proof mass; this wafer is anodically bonded to two capping wafers on both sides. Instead of diffusing piezoresistors on the suspension cantilevers, capacitor plates are defined on both sides of the proof mass and are aligned with the capacitor plates on the capping wafers to form two differential capacitors.

In one design, the proof mass is symmetrically suspended by eight cantilever beams located on both sides of the wafer. The suspension beams are aligned with the diagonal axes of the proof mass to obtain maximum symmetry, and to allow ease in undercutting the beams with the anisotropic etchant used to form the proof mass. An ASIC chip using switched capacitor techniques is used to amplify the signal.

Bulk micromachined accelerometers are typically a few millimeters on each side. Anodic bonding can be used to form the fixed electrodes and to provide overrange protections. Because the capacitance is determined by the gaps of three bonded wafers, the anodic bonding step is crucial in determining the yield of the accelerometers. The accelerometer size is important because it is directly related to its cost.

There has been intensive developmental work in the field of bulk micromachined capacitive accelerometers in both the U.S. and Europe. CSEM of Neuchâtel, Switzerland, has designed a unit with a 1-μg resolution for inertial navigation applications at 1 Hz.[33] The device (see Fig. 7.7) is fabricated by means of conventional anisotropic etching of bulk silicon from both sides so that the two cantilever suspensions are aligned with the center of the proof mass to produce a low cross-axis sensitivity of 0.4%.

In the U.S., Triton Technologies, Inc., developed an accelerometer with a resolution of submicrograms. The basic fabrication process for the Triton/IC Sensors

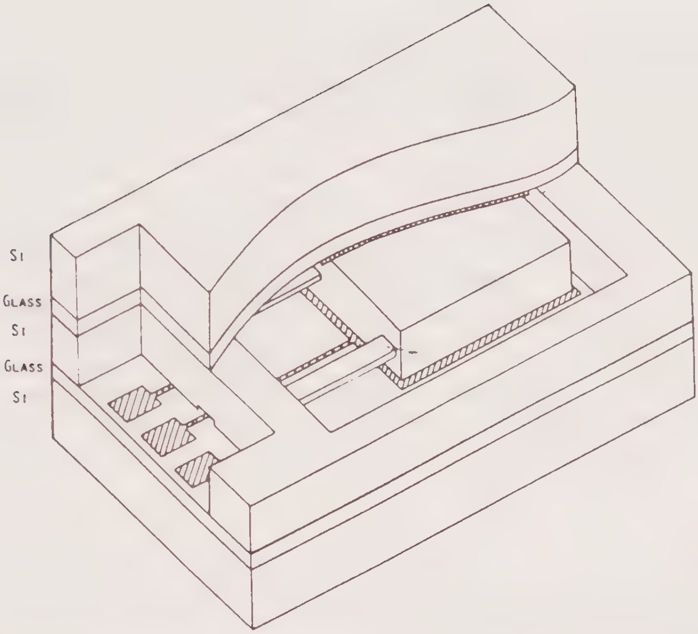


FIG. 7.7 Schematic view of a bulk-micromachined, capacitive accelerometer.[33] (© Elsevier Science)

accelerometer in Fig. 7.8 entails the double-sided etching of one wafer to form the proof mass and the folded, symmetrical suspension.[34] The proof mass is bonded to the upper and lower capping wafers to define the sensing capacitance and overrange protection. This design releases most of the process-induced, thermal, and packaging-induced stresses. The interconnection lines from the sensing capacitance to the bonding pads are shielded with driven guard electrodes on top and bottom to minimize parasitic capacitance and prevent coupling.

As shown in Fig. 7.9, the Triton device first passes the sensing signal through a unity gain buffer that drives the active guard electrodes. The lead-lag block ensures loop stability. A 1-bit quantizer converts the analog signal into a high-frequency, density-modulated serial bitstream representation of the input acceleration The clock frequency is much higher than the Nyquist sampling frequency (1024×). The quantization noise is shaped so that the low-frequency baseband noise is significantly reduced and the high-frequency noise is filtered out by the decimation filter. With this modulation technique, high S/N ratio and large dynamic range can be achieved without need for precision components. This accelerometer is reported to have 0.1 mg resolution and 120 dB dynamic range. Its cross-axis sensitivity is described as <0.001%. In addition to the complexity of the interface electronics, fabrication of the sensing element in bulk micromachined, capacitive accelerometers can be very complicated. In Triton's design, 27 lithography steps are taken on 5 wafer surfaces of the 3 bonded wafers; cost is directly related to the system's complexity.

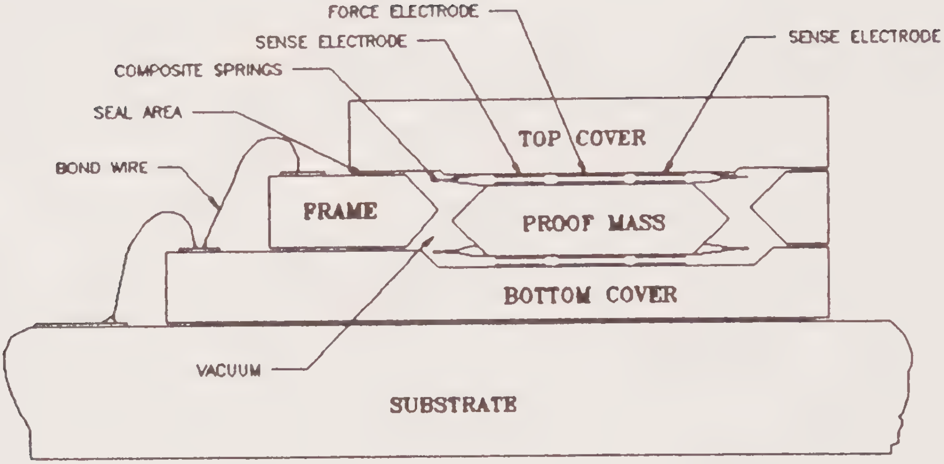

FORCE ELECTRODE
SENSE ELECTRODE
SENSE ELECTRODE
COMPOSITE SPRINGS
SEAL AREA
BOND WIRE
TOP COVER
FRAME
PROOF MASS
BOTTOM COVER
VACUUM
SUBSTRATE

FIG. 7.8 Cross-sectional view of a force-balanced, capacitive accelerometer.[30] (©1990 IEEE)

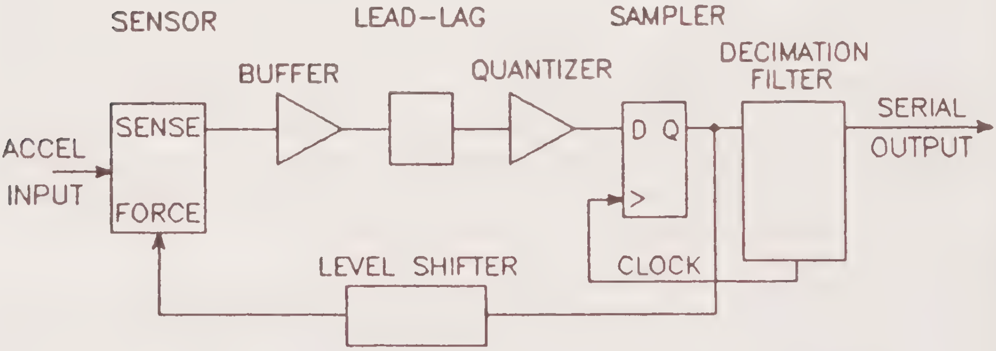

SENSOR
LEAD-LAG
SAMPLER
BUFFER
QUANTIZER
DECIMATION FILTER
ACCEL INPUT
SENSE
FORCE
D Q
>
SERIAL OUTPUT
LEVEL SHIFTER
CLOCK

FIG. 7.9 Block diagram of the sensing electronics for a force-balanced, capacitive accelerometer.[30] (©1990 IEEE)

7.5.1.1 Surface-micromachined capacitive accelerometers

Surface micromachining is a processing technique developed in the past decade. It uses low-pressure chemical vapor deposition (LPCVD) or electroplating to deposit and define the proof mass and its suspensions on the front side of the silicon wafer. A sacrificial spacing layer, typically LPCVD oxide, is removed after the structural film is deposited to release the structure from the substrate. No anodic bonding is required. The sensors are typically a few hundred microns on each side, with the thickness of the mechanical and spacer layers determined by the deposition times of those layers. The performance of these devices depends critically on the uniformity and mechanical properties of these deposited layers, including the effects of residual stress in the films. Because the sensing element is small and on only one side of the wafer, interface electronics are often integrated on the same chip.

Motorola has reported a 50-g capacitive accelerometer consisting of three polysilicon layers; the first and third form the upper and lower differential capacitor plates and the second forms the suspended proof mass.[35] Designed for automotive air bag applications, the device features a twin-cantilevered suspension. Limit stops are fabricated from the third polysilicon layer, which overlaps extensions of the suspension cantilevers. A switched capacitor readout circuitry is fabricated on a separate CMOS ASIC chip and packaged adjacent to the sensor chip. The accelerometer has an open-loop sensitivity of 40 mV/g and a bandwidth from 0.6 Hz to 1 kHz.

7.5.2 Force-Balanced Accelerometers

Sensitivity and dynamic range are important accelerometer specifications. Sensitivity is limited by the stiffness of the mechanical suspension, with a compliant suspension desirable for high sensitivity. In an open-loop design, the dynamic range is constrained by the nonlinearity of the suspension, and the transduction by piezoresistors or capacitors. A stiffer (less compliant) suspension improves the dynamic range by limiting the displacement. Trade-offs therefore must be made between sensitivity and dynamic range for open-loop accelerometers.

In force-balance accelerometers, the suspended proof mass deflects in response to the acceleration. The signal (strain or displacement) is sensed, amplified, and fed back to counteract the displacement of the mass due to the input acceleration. During operation, the structure is virtually stationary and the linearity and effective spring constant are determined by the electronic control. Because the suspension can therefore be made very compliant without sacrificing dynamic range, and the accelerometer can be made ultrasensitive, these devices are commonly used for navigational applications. The piezoresistive sensing technique is rarely used in force-balance accelerometers due to its large temperature sensitivity; capacitive sensing is most frequently selected due to process compatibility. In conventional, nonmicromachined force balance accelerometers, capacitive sensing and magnetic feedback are often used to avoid feedthrough. Due to the incompatibility of magnetic materials with the IC process, magnetic feedback is very costly and rarely used in silicon-based accelerometers.

The fabrication of the sensing element is essentially the same as that for open-loop accelerometers, but sophisticated electronics must be designed for the feedback control. The resulting cost increase can be reduced by integrating the electronics on the same chip with the sensing element.

In surface micromachined, force-balance capacitive accelerometers, the fabrication is simpler because only one wafer surface is processed. This and the small size of the sensing element make the integration of the sensing and control electronics on the same chip very attractive. Analog Devices, Inc., has reported a major step toward integrating the sensing element and the sensing, control, and interface electronics.[36] This accelerometer is fabricated by means of a polysilicon microstructure process, where polysilicon is deposited and patterned on a sacrificial oxide and later freed by etching away the sacrificial oxide with hydrofluoric acid. As shown in Fig. 7.10, the moving plates are suspended by the four tethers, and the fixed comb fingers sense capacitance change as the plate displaces.[37] The second polysilicon layer forms the movable proof mass, the interdigitated fingers for levitation, and the folded suspensions, which are designed to ensure a low cross-axis sensitivity.

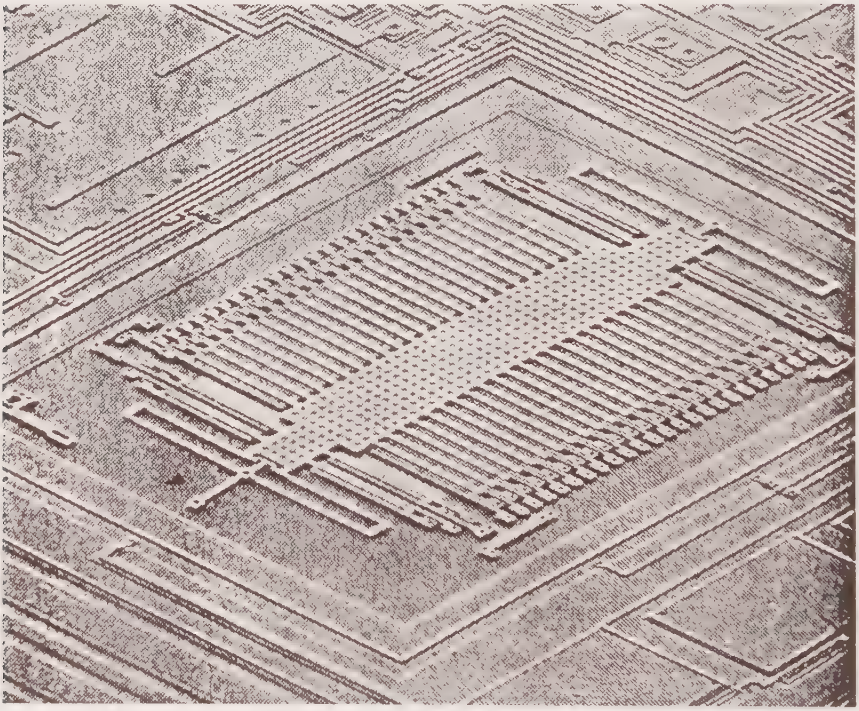


FIG. 7.10 SEM photograph of a surface-micromachined accelerometer using suspended polysilicon elements.[37] (©1995 IEEE)

Sigma-delta modulation, popular in A/D converter designs, has several advantages that make the technique attractive for feedback control of accelerometers. First, the feedback system has intrinsic linearity and a large dynamic range because the feedback is a 1-bit pulse train and the feedback force is proportional to the density of the pulse train. Second, the requirements for the readout circuits are less demanding because the 1-bit quantizer concerns only the polarity of the signal. Third, its direct digital output facilitates interfacing with the rest of the system. The whole chip, including the sensing element and the readout electronics, is only 2.5 by 2.5 mm. In force-balance accelerometers, the mechanical properties are dominated by the electronic feedback controls. Performance is therefore strongly influenced by the electronic system characteristics, which are more controllable and reliable than the mechanical properties. Competence in sensor, integrated circuit (IC), and system design is required in order to optimize the overall performance.

7.5.3 Resonant Accelerometers

Solid-state resonant accelerometers originated in the 1970s with research conducted by Kearfott Guidance and Control on devices using push-pull quartz resonant strain gauges. Sundstrand Data Control has recently developed a family of quartz resonant accelerometers that use double-ended tuning forks as the vibrating strain gauges rather than clamped-clamped beams. General Motors Research Laboratories has reported a two-axis resonant microaccelerometer.[38] This device measures the component of acceleration in the plane of the substrate. The inertial load on the proof mass causes differential axial loads in the two pairs of polysilicon resonant-microbridge strain gauges, shifting their frequencies. The component of acceleration that is aligned with a pair of microbridges is measured by the frequency difference between the pair. Acceleration components normal to the chip produce no signal because the vertical displacement of the proof mass has no first-order effect on the resonant frequencies of microbridges and, further, any effect appears as a common signal and is canceled in the differential output. The constraining bridges are fabricated on the back side of the device in order to prevent excessive vertical displacement that would cause the proof mass to contact the limit stop. This device has a sensitivity of 160 Hz/g. Capacitive excitation and detection drives and senses the polysilicon resonant microbridges.

Resonant accelerometers present the microsensor designer with exacting demands. For example, in order to minimize mode jumping, the vibration mode of the strain gauges should have high sensitivity to compressive or tensile axial loads and yet not be close to other modes. In addition, a high quality factor for the resonator is desirable, implying that a low-pressure or vacuum ambient is optimum. The proof mass and suspensions, however, should be critically damped to avoid ringing. One solution would be to encapsulate the resonators in a thin film vacuum shell. As a final note, the relatively stiff suspensions of this class of accelerometers make package-induced stresses a problem in compensating the packaged device.

7.5.4 Tunneling Accelerometers

Displacement transducers based on electron tunneling have been proposed for use in a wide variety of physical sensors. These sensors utilize the high position sensitivity of electron tunneling. Displacement resolutions approaching 10^{-4} Å/√Hz have been obtained by Waltman and Kaiser[39] and by Kenny et al.[40] for tunnel transducers. Employing a gold tip formed by breaking a gold wire, and a piezoelectric bimorph as both the proof mass and the control element of the system, Waltman and Kaiser observed an acceleration resolution of 10^{-5}g/√Hz with a bandwidth of 3 kHz. Baski et al.[41] utilized a beryllium-copper spring and a 2-g proof mass mounted within a miniature scanning tunneling microscope (STM) with a tungsten tip and a graphite counter electrode to measure acceleration resolution better than 10^{-4} g/√Hz in a 200-Hz band.

Sensitive compact accelerometers capable of detecting accelerations on the order of 10^{-7} to 10^{-9} g/√Hz in a band from several hertz to several kilohertz appear to be possible. Such accelerometers can have application as sensitive acoustic sensors. Growing needs for small, low-cost accelerometers include network applications using an array of distributed units. Accelerometers function by employing proof masses whose motion responds to accelerations. Motion of the proof mass relative to the accelerometer case is measured directly; the motion of the case is determined from the motion of the mass and knowledge of the sensor dynamic response.

Recent work has focused on the development of tunneling accelerometers based on silicon micromachining. The tunnel accelerometer described by Kenny et al.[42] consists of a single suspended mass component, the proof mass, a folded-cantilever proof mass suspension, and a gold-coated silicon tip. Electron tunneling takes place between a tip on the proof mass and a fixed counter-electrode on the case. Operation of this accelerometer uses electronic feedback circuitry to maintain a constant tip-to-counter electrode spacing, as in scanning tunneling microscopy, by controlling an electrostatic deflection voltage on the suspended mass. When the sensor is accelerated, the proof mass experiences an inertial force which causes its motion to lag that of the sensor. During operation of this device, acceleration forces on the cantilever are balanced by electrostatic forces generated by a feedback circuit, so as to keep the cantilever in the same relative position with respect to the frame. In order to achieve stable operation, the bandwidth of the feedback circuit is restricted to frequencies below the natural frequency of the proof mass, 200 Hz in this case.[43] An acceleration resolution of 10^{-7} g/√Hz at 10 Hz is inferred from noise measurements.[44] A new tunnel accelerometer has been designed to achieve high resolution over a broad band of several kilohertz. Broadband operation is achieved in this case by use of an innovative dual-cantilever structure.[45]

7.5.4.1 Alternative tunneling structures

Yao and co-workers have developed a series of high-resolution silicon microfabrication techniques for use in construction of two-dimensional scanning/tunneling structures.[46] These devices achieve very high resonant frequencies and a capability for integration of electronics on the microstructure through use of lateral oxidation and etching to release single-crystal micromechanical elements from the substrate. Kobayashi et al.[47] have demonstrated a surface-micromachined lateral tunneling unit, which is fabricated by deposition of sacrificial layers and polysilicon, with subsequent patterning, etching and metallization. Tunneling currents were observed, but with weaker than expected variation with applied actuator voltage.

7.5.5 Conclusions

Silicon microfabricated accelerometers have already achieved commercial success. What is not completely clear at present is the match between the spectrum of accelerometer applications and the various designs and fabrication technologies that have been demonstrated recently. For example, open-loop piezoresistive accelerometers are simple and can be made robust, but they have high temperature sensitivity. Capacitive accelerometers are superior in this respect, although the fabrication process (for bulk-micromachined designs) is generally more involved. High-precision silicon accelerometers will almost certainly be capacitive. For capacitive accelerometers, the decision to close the loop is an option for the system designer, since the basic structure of the sensing element is not affected by the use of analog or digital feedback to the capacitive plates. Although the technology is less developed than the force-balance approach, the use of resonant strain gauges is a competitive technique for high-precision sensing.

Bulk micromachining, supplemented by wafer bonding, has been the dominant fabrication technology. Surface micromachining, however, has the advantages of single-sided processing, smaller die size, and integration of the sensing, control, and interface electronics. Thin film microstructures have the disadvantage of small proof masses and a nearly 2-D structure. It remains to be seen whether surface micromachined sensing elements, combined with sophisticated on-chip signal processing electronics, can achieve high sensitivity and large dynamic range for high-performance applications.

7.6 FLOW SENSORS

7.6.1 Introduction

Advances in thin-film deposition techniques and etching methods for silicon microstructures have made possible a technology for the miniaturization of thermal devices, including fluid flow sensors. The advantages in using micromachined devices in flow sensing applications include fast response, low power consumption, and the measurement of small flows. However, as the size of the sensors is decreased, their interactions with the thermal and fluid environment can be influenced by small-size effects. The magnitude of these effects can be estimated from the evaluation of the following nondimensional parameters[48]: Reynolds number $\mathrm{Re}=Ud/u$; Knudsen number, $\mathrm{Kn}=l_m/d$; and the Fourier number, $\mathrm{F}o=d^2/\alpha t$, where U is the characteristic fluid velocity, d is the characteristic dimension, u is the kinematic fluid viscosity, l_m is the mean free path length of a gas molecule, α is the thermal diffusivity, and t is the time.

Given a structure dimension on the order of 10 μm, which is typical of a microsensor device, air velocities less than 1 m/s or water velocities less than 0.1 m/s would represent Reynolds numbers less than 1. Thus, the fluid environment would be laminar at that level and possibly even in the Stokes flow regime for smaller velocities. For air pressures and temperatures near standard conditions, the mean free path l_m is approximately 6.5×10^{-8} m. Thus, Knudsen numbers based on a gap thickness of 1 μm would be on the order of 0.1. This value would correspond to the transition between continuum gas conditions and free molecular flow. For higher temperatures, lower pressures, smaller gas molecules, or smaller dimensions, the Knudsen number would increase, thus forcing the flow regime more toward a free molecular flow.

A low-pressure sensor described by van Herwaarden and Sarro[49] exploited the Knudsen effect. This device consisted of a membrane suspended between two heat sinks. The gas to be measured filled the gaps on either side of the membrane. The membrane was heated resistively, and the temperature differences between the membrane and the heat sinks were measured with thermopiles. The measured power input to the membrane was used to measure the overall thermal conductance as a function of ambient pressure. The transition from gas continuum transport to molecular heat transport was observed at a pressure corresponding to a Knudsen number of 0.1. This use of small-size effect physics provided a fairly sensitive, low-pressure device.

The Fourier number is of interest in assessing the response time of temperature sensors. If the sensor material is fused silica of a dimension of 10 μm, Fourier numbers of the order of 1 are represented by time scales on the order of 300 μs. Steady-state temperature profiles within a surrounding gas cavity of 1 μm dimensions will occur four orders of magnitude faster. Thus, one primary advantage of the use of microsensors for thermal measurements is their remarkably fast response times.

7.6.2 Gas Flow Sensors

Many kinds of gas flow sensors have been developed.[50] Two aspects which are crucial for high-accuracy gas flow measurements include not only a well-characterized and reproducible transduction mechanism, but also a well- defined and repeatable gas flow geometry. The transduction mechanism itself is important because it plays a major role in determining resolution, linearity, accuracy, speed, and other performance characteristics. The geometry is equally important, however, since the flow rate is proportional to the cube of the effective radius through which the gas is flowing, and small variations in channel dimension may result in large gas flow measurement errors. In addition, the size and shape of the flow channels and the entrance and exit geometries influence characteristics such as laminar versus turbulent flow, which in turn affect heat transfer parameters.

Silicon micromechanics has been exploited to provide both these functions. Bridges have been designed with unique shapes to optimize strength, minimize thermal losses to the supporting substrate, maintain structural stability under high gaseous velocities, provide very high corrosion resistance, and simplify fabrication procedures. The thermal resistance from the bridge to the substrate has been measured to be greater than 10,000 C/W. This makes it possible to operate the thin film resistive element at greatly reduced power while still generating substantial temperature differences and providing response times in the millisecond range.

An example of a thermal mass flow sensor is illustrated in Fig. 7.11.[51] A pair of titanium microheaters is sandwiched between two SiO_2 thin membranes. To achieve thermal isolation, the heaters are suspended with four small silicon islands on a Pyrex plate. This sensor exhibited high sensitivity for very small flows from 0.1 μl/min.

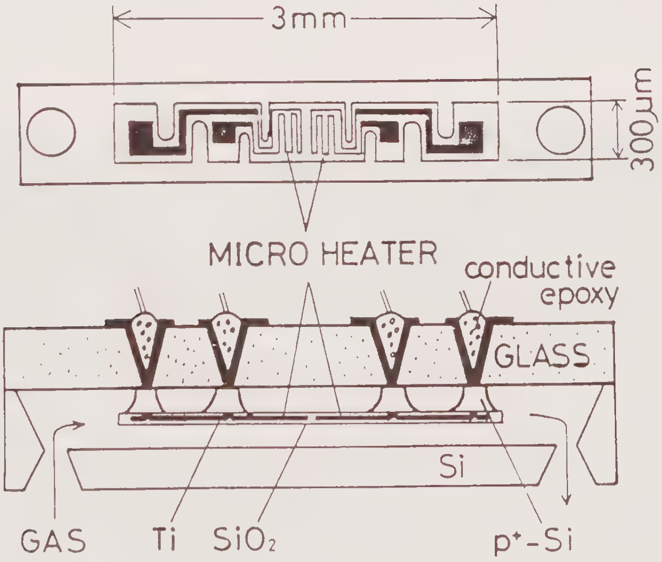


FIG. 7.11 Cross section of a micromachined flow sensor using suspended Ti heater elements.[51] (©1991 IEEE)

Indirect methods of flow sensing have also been integrated into microsensors. Anemometry is a common flow detection technique applicable to small-scale devices. One device produced acceptable performance for air flow in the range of

0–2 m/s using a polysilicon-bridge anemometer positioned 3 μm from the substrate. The sensor output consisted of a voltage measurement across a temperature-dependent resistor carrying a constant 300 mA. Voltage signals on the order of 30 mV were obtained when the sensor was exposed to an air velocity of 1 m/s. Since the output voltage is essentially proportional to the local heat transfer coefficient, it varies with the square root of the Reynolds number and the response would not be linear with velocity. Voltage proportional to the square root of the free-stream velocity was observed with this sensor.

A gas flow sensor based on the thermoelectric effect can be made by using the aluminum/polysilicon contacts provided by the CMOS process.[52] An oxide beam contains an integrated thermopile with a polysilicon heater and the hot contacts at the tip (Fig. 7.12), while the cold contacts are placed on the bulk silicon. The small Seebeck coefficient of the aluminum/polysilicon thermocouple is compensated by the excellent thermal insulation provided by the beam; appreciable heat conduction is provided only by the metal and polysilicon lines connecting the contacts. Sensitivities of 28 V/W and 1 mV/K were observed for the whole thermopile. With laminar nitrogen flow, such a sensor produces an output voltage of 9 mV/ $(m/s)^{0.5}$ for 5 mW heating power and hence 1.8 $V(m/s)^{0.5}/W$ sensitivity.

Similar devices for the measurement of flow velocity have been fabricated with the heating element mounted directly on a glass substrate but do not exhibit the high

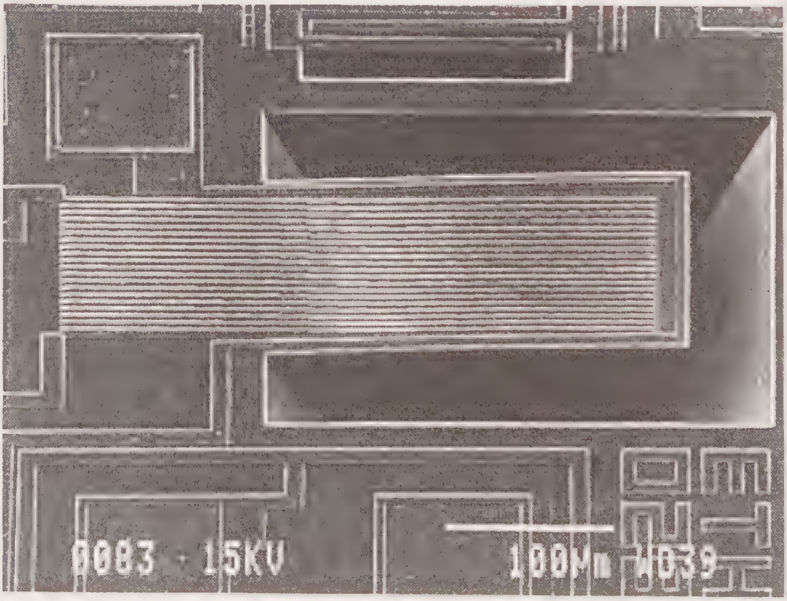


FIG. 7.12 Gas flow sensor fabricated using CMOS technology and aluminum/polysilicon thermopile temperature sensors.[52] © (Elsevier Science)

sensitivity obtained by locating the heating and detection elements on bridges suspended above the substrate. Nevertheless, the simplicity of the design with the heaters and thermistors fabricated directly on the surface may yield better survival in harsh environments and demand less recalibration. These devices have also been shown to be amenable to multielement design, allowing simultaneous detection of flow velocity and direction.

7.6.3 Surface Shear Sensors

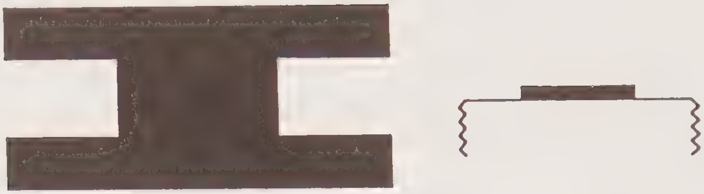
The direct measurement of surface shear in turbulent flow is possible with a micromachined floating-element sensor.[53] This device, shown in Fig. 7.13, allows for the measurement of surface shear stress in a turbulent boundary layer by detecting the floating-element deflection through a change in capacitance. The tethers behave as linear springs, thus providing a direct relationship between surface shear and displacement. The 500 μm × 500 μm element can measure shear stresses of 1 Pa. It is expected that this device should have a bandwidth of about 20 kHz as opposed to 100 Hz for conventional, large-mass floating elements, thus allowing a more detailed examination of the small-time-scale events that are characteristic of highly turbulent flow.

7.6.4 Humidity Sensors

The measurement of humidity has conventionally used the difference in temperature between wet and dry bulb thermometers as a measure of the water content of the air. In a silicon microstructure, the change in electrical conductivity of a layer as a function of absorbed water, or the change in the permittivity of a material with respect to surface moisture, can be used to measure humidity.[54] Changes in the relative humidity produce a linear change in capacitance in an adsorbing polyimide thin film. For film thicknesses on the order or 5 μm, differences in capacitance of about 4 pF were measured over the full range of relative humidities. The reported response times were relatively slow (approximately 1 min), primarily due to the rate-limiting processes of mass transfer within the adsorbing polyimide film.

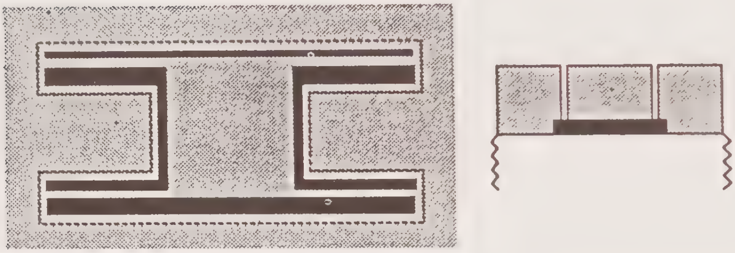
7.6.5 Liquid Flow Sensors

The measurement of small liquid flows is becoming increasingly necessary in analytical applications, such as flow injection analysis, and in medical applications, such as drug delivery. The flow rates in analytical systems are in the range of 0.1–1 ml/min, and high accuracy and fast dynamic response are of prime importance. The flow rate in medical applications can be as low as 0.5 ml/hr (0.008 ml/min) for ambulatory infusion applications.



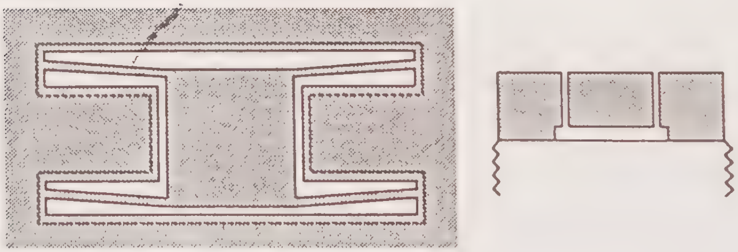
DEPOSIT AND PATTERN SPACER MATERIAL

(a)



DEPOSIT AND PATTERN POLYIMIDE

(b)



WET ETCH SPACER MATERIAL
ELEMENT SHOWN DEFLECTED UNDER SHEAR-STRESS LOAD

(c)

FIG. 7.13 Process flow for a surface-micromachined shear-stress sensor. The cross sections shown on the right are through the center of the sensing element.[53] (©1988 IEEE)

7.6.5.1 Thermal dilution flow sensor

Thermal methods have typically been applied to the measurement of liquid flow rate. The thermal dilution principle exploits the proportionality between flow and the extent of cooling of an immersed, self-heating probe to detect flow rate. A thermal dilution flow sensor is illustrated in Fig. 7.14.[55] It consists of three resistors located in the middle of a flow channel. Current is passed through the middle resistor, generating heat. The heat distribution from the middle resistor is sensed with two temperature-sensitive resistors located symmetrically up and downstream. The temperature difference between the upstream and downstream sensors is converted using bridge circuitry. A sensitivity on the order of 25 μV/K for a heater dissipation of 10 mW was obtained using isopropanol as a liquid.

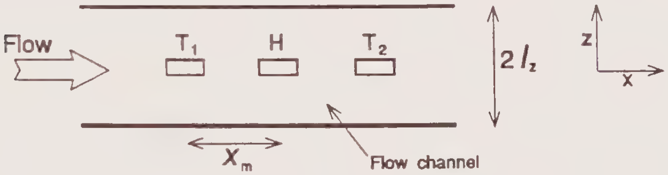


FIG. 7.14 Cross section of a liquid flow sensor. A flow through the sensor produces a temperature difference between the upstream and downstream sensors resulting from the transport of heat from the central heater.[55] (©1993 IEEE)

7.6.5.2 Thermal transit-time sensor

A small-sized thermal transit-time flow sensor, using square heat waves as tracers, has been developed for measuring liquid flows as small as 0.1 ml/min in microchannels.[56] When the flow sensor operates in a thermal transit-time mode, the heating diode is supplied with a square-wave voltage at an appropriate frequency to inject heat pulses as tracers into the fluid. The square heat pulses are transferred by the axial thermal transport in the flow channel, causing periodic temperature fluctuations downstream, which are detected by a measurement diode. In steady state, the phase shift of the first harmonic of the temperature fluctuations with respect to the square waves will be inversely proportional to the mean velocity of the fluid.[57]

This principle has little dependence on the ambient temperature and on the properties of the fluid, and the transit time can be determined accurately. This type of flow sensor demonstrates high accuracy, better than 0.2% (at the flow rate of 0.2 ml/min), and good long-term stability. In addition, these flow sensor have demonstrated excellent linearity.

7.6.6 Differential Pressure Flow Sensors

The structure of a differential pressure flow sensor is shown in Fig. 7.15.[58] In laminar flow conditions, with low Reynolds numbers, the pressure difference across an orifice is proportional to the flow rate. The differential pressure is detected in this case by the deformation of a cantilever having a piezoresistor stress sensor. A similar type of microflow sensor, fabricated using a buried piezoresistor stress sensor, was reported by Nishimoto et al.[59]

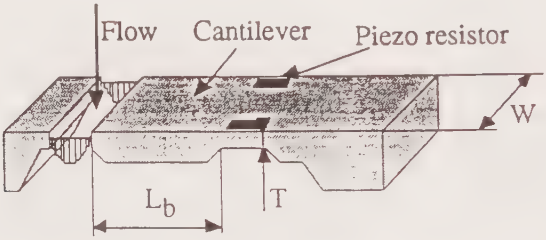


FIG. 7.15 Flow sensor based on the principle of drag force. The cantilever is 1 mm long and the effective diameter of the flow channel is 58 μm.[58] (©1993 IEEE)

Other flow sensing devices have been reported, including one by Suminto et al.,[60] who created a venturi-tube, differential-pressure liquid-flow sensor. It consists of a Pyrex tube and two pressure sensors. Also, Carr developed a fiber optic liquid-flow sensor based on photon correlation spectroscopy.[61]

7.6.7 Conclusions

The use of micromachining technology for flow sensors for very low gas and liquid flows appears very promising from the early development work, but no breakthrough commercial application has appeared to speed the progress of these devices from prototypes to production. In large part this can be attributed to the lack of conventional methods to measure these low flows and therefore no established market to penetrate with lower cost micromachined devices.

7.7 ACTUATORS

7.7.1 Introduction

There are two general classes of actuators for micromachined devices, those that are fabricated separately from the devices (external) and those fabricated as part of the devices (integrated). Historically, integrating powerful actuators has been difficult; therefore many researchers have chosen to use external actuators in order to achieve the desired results. The use of external actuators is likely to be more expensive for commercial applications, as the individual actuators need to be assembled with the micromachined elements, but this may be acceptable in some applications.

The external actuators that have typically been used include electromagnetic actuation, either by direct actuation with a solenoid plunger, or using an integrated region of permeable magnetic material in conjunction with an externally generated magnetic field. Piezoelectric drivers, typically disks or bimorphs, have been used in many cases, usually by adhesively attaching the piezoelectric devices to a micromachined device. External pneumatic systems can be used to drive miniature actuators, since it is relatively easy to attach small diameter tubes to miniature devices. More recently, external shape memory alloy (SMA) actuators have been coupled to micromachined devices. Integrated actuators have typically used electrostatic, electromagnetic, thermopneumatic, bimetallic, or the shape memory effect to provide actuation force.

Historically there have been two main applications that have driven the development of micromachined actuators: valves and optical scanners. The use of external and integrated actuators will first be described in relation to their use in valves. The creation of a fully integrated, batch-fabricated, silicon valve has been a goal for more than two decades. Early efforts at Stanford University in the 1970s resulted in two valve designs, one pneumatic and one utilizing a standard magnetic solenoid for the actuator, for use in a miniature gas chromatography system.[62, 63] While these valves had many significant features, they were relatively expensive, difficult to assemble, and were not capable of being batch fabricated. The pneumatic valve used a silicon diaphragm to close against an etched silicon valve seat, resulting in a valve that would often leak more than could be tolerated for a critical chromatographic injection valve.

Conventional valves for pressure and flow control have typically used magnetic actuation in the form of solenoids or motors to drive diaphragms or spool valves. While miniature planar solenoids have been made using micromachining techniques, the magnetic fields developed have not approached the values commonly achieved in conventionally wound solenoids and do not provide sufficient force for most valve applications. For very small structures, electrostatic actuation becomes more attractive, but achieving high pressure valves with electrostatic activation has proved difficult. Actuation is possible using piezoelectric drivers.[64] The force and deflection properties of materials such as PZT produce high forces, but with very small deflection for button-type actuators, and the complicated assembly and high voltage operation of the actuators are unattractive for many commercial applications.

For very small structures, thermal actuation is attractive due to the geometrical advantages of scaling. The amount of thermal mass decreases as the volume of heated material decreases; the thermal loss decreases as the support thicknesses decrease; while the force per unit area remains high. There are a variety of general mechanisms that can be used as thermal actuators, all relying on either the linear or volume expansion of a material. Thermal actuators of many types have been used in the automotive industry for some time, typically as drivers for latches or indicators. One such device uses a wax material that shows a large volume expansion upon melting and another, a single-stroke heat engine, uses the vapor pressure of a superheated liquid to generate pressures on the order of 250 psi.[65] This principle was used by Zdeblick, initially at Stanford University, for a miniature thermally activated valve using the vapor pressure generated in a heated fluid.[66]

The performances of active microvalves and micropumps depend strongly on the performance of the actuator. In fact, the size of the devices is largely determined by that of the actuators, both for external and integrated actuation. The actuators used so far are classified into two categories: miniaturized conventional actuators and micromachinable actuators. Many kinds of actuators have been studied and used. Since the size of the active microflow control devices is actually restricted by that of the actuator, it is very important to miniaturize the actuator.

7.7.2 Microvalve Examples

Microvalves developed so far can be classified in two categories: active microvalves (with an actuator) and passive microvalves or check valves (without an actuator). Active valves can be further classified as having either external actuators or integrated actuators.

7.7.2.1 Microvalves with external actuators

Solenoid plunger. Electromagnetic actuation can be realized by using a solenoid plunger. The force developed by these actuators depends on the applied current and the number of turns of wire in the solenoid. External solenoid actuators were used with good success on an early version of the integrated gas chromatograph.[67] The solenoid-drive valve in a later version of this design used a Teflon™-coated Kapton™ diaphragm that was mated to an etched silicon valve seat. Here a spring provided the normally closed force to the diaphragm, and the solenoid was used to remove the spring pressure, allowing the valve to open. Typical cycle times for this valve were on the order of 2 ms and the on-to-off flow ratio was over 1 million to one, which was necessary for this application. Integration of such a valve is, however, difficult because of the difficulty in obtaining sufficient turns in a solenoid and in providing a low-loss magnetic return path. Alignment and attachment of the actuator to the substrate generally requires more area than the active valve area.

Piezoelectric actuation. Piezoelectric actuators are commercially available as both disks and cantilever beams. These actuators are usually used as bimorphs, where two layers of piezoelectric material are stacked so that the applied voltage causes one layer to contract and one to expand, providing for more efficient response. The deflection from bimorph cantilever beams can be relatively great, but the corresponding force is low. The displacement available from a stack of piezoelectric disks can be quite large, but the stroke is small, making precision assembly necessary. For example, a pressure of 350 kgf/cm^2 at 150 V applied voltage has been generated by stacking 64 thin ceramic plates with a total size of 1.4 × 3 × 9 mm.[68] This stack has a response time of about 0.1 ms, but the maximum stroke is only 8 μm.

A normally closed microvalve using this stack of piezoelectric actuators is illustrated in Fig. 7.16. It has a glass-silicon structure with a movable diaphragm. To avoid leakage, a free-standing nickel gasket was used. A three-way microvalve was fabricated by modifying this structure as shown in Fig. 7.17.[69] A microvalve having large stroke for liquid flow control was also fabricated using a bimorph-type piezoelectric actuator.

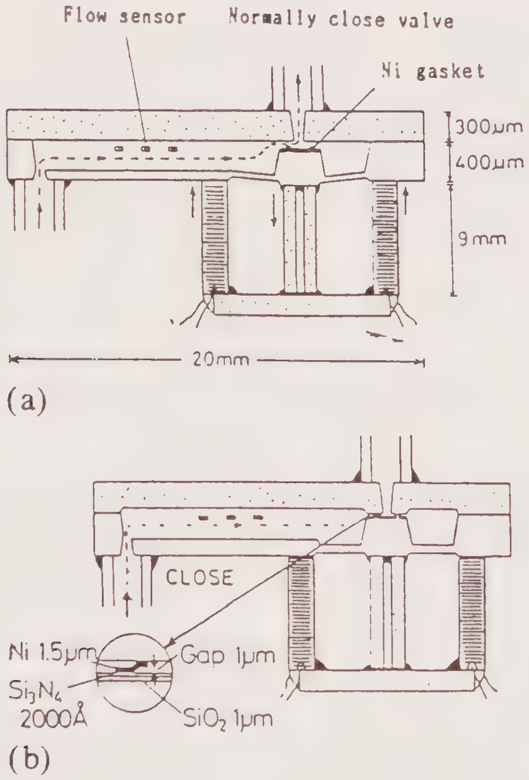


FIG. 7.16 Cross section of a normally closed microvalve driven with stacked, piezoelectric elements. When the actuators are driven, they pull the center boss away from the valve seat, opening the valve.[69] (© Elsevier Science)

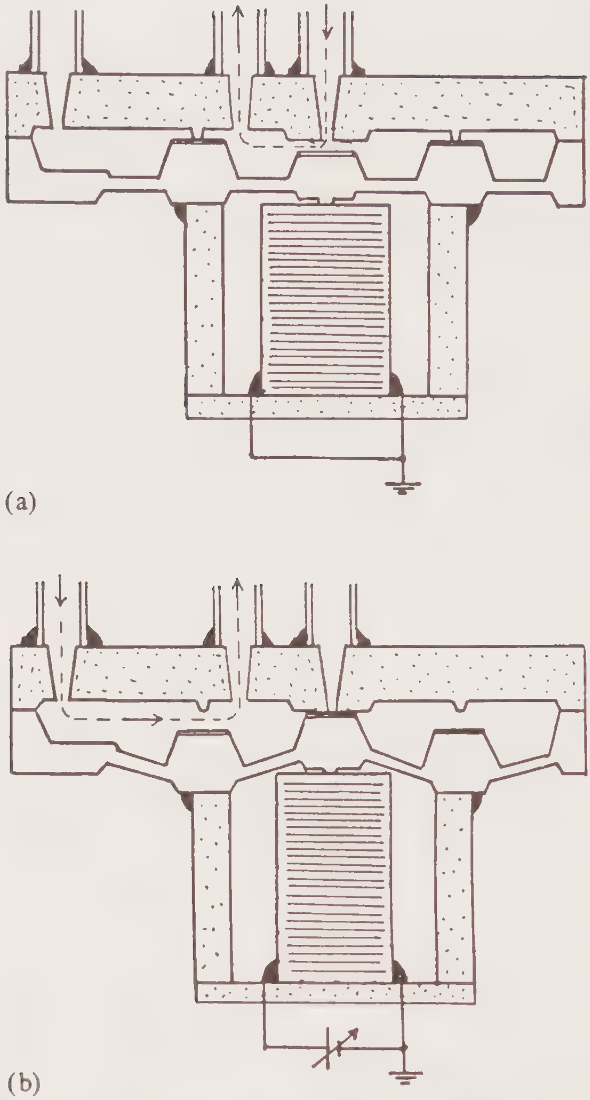


FIG. 7.17 Cross section and operating principle of a three-way microvalve driven with piezoelectric elements.[69] (© Elsevier Science)

Pneumatic actuation. An external pneumatic system can be used to drive actuators. Such a system typically consists of a commercially available electromagnetic valve and a source of compressed air. This type of valve was used on the first versions of the integrated gas chromatography system.[70] The displacement and generated force can be controlled over a wide range. In this case, the response time depends mainly on the flow conductance of the tubing connecting the external valve to the micromachined elements. The size of the micromachined components can be quite small, but miniaturizing of the external valves is difficult. It is possible in some cases to have a small number of external valves driving a large number of micromachined devices, making a reasonable system configuration.

Shape memory alloy external actuators. Small actuators can be made by using a shape memory alloy spring working against a bias spring. The SMA coil returns to its original shape when it is heated above its critical temperature. With a Ti-Ni wire of 0.5-mm diameter wound as a coil with a 3 mm diameter, a pressure of about 0.2 MPa (30 psi), and a displacement of about 1 mm can be obtained.[71] The critical temperature of the SMA in this case is around 50 °C and can be achieved by using the SMA wire itself as a heater. To obtain a reasonably fast response time, a current on the order of 1 A is required. It is difficult to control the displacement precisely in these devices, so they can only be used as on-off valves.

7.7.2.2 Microvalves with integrated actuators

Electrostatic. An electrostatic actuator in its simplest form has a movable planar electrode and a fixed electrode. The electrostatically generated pressure P_{el} can be calculated by

$$P_{\mathrm{el}} = \frac{1}{2}\varepsilon_0\left(\frac{V}{d}\right)^2 , \tag{7.7}$$

where ϵ is the dielectric constant of the material between the plates, d is the gap between the plates, and V is the applied voltage. The generated force of the actuator is inversely proportional to the second power of the distance. For example, with an applied voltage of 10 V, a gap of 10 μm, with air between the plates (ϵ_0=8.86 × 10^{-12} F/m), the electrostatic pressure is only 4.4 Pa. If the voltage is increased to 100 V and the gap decreased to 5 μm, then the pressure is increased by a factor of 400, but to only 1800 Pa (0.25 psi). Thus it is difficult to generate both a large electrostatic pressure and large stroke. The nonlinear force vs. deflection characteristics of this actuator makes it susceptible to snap-in, where the electrostatic force overwhelms a linear restoring (spring) force, when the gap is about one-third of the initial gap.

A simple cantilever-type microvalve, shown in Fig. 7.18, was fabricated by surface micromachining.[72] The force generated by this valve is small when the electrode gap is large. The controllable pressure range is also limited, with a pressure over 14,000 Pa (2 psi) causing the valve to open even when the voltage is applied.

In order to provide enough force to actuate a miniature valve, the pressure of the fluid can be used to balance the force on the moving parts of the valve. The schematic cross section and basic concept are illustrated in Fig. 7.19.[73] This valve was made of stacked silicon structures assembled by fusion bonding. The fluid pressure produces both an upward force to close the valve and a downward force to open it. The force required to actuate the valve can be only a small fraction of the total pressure of the fluid when the valve has properly designed top and bottom geometries. This valve was tested both pneumatically and electrostatically, and provides a small flow conductance because of the reduced valve stroke.

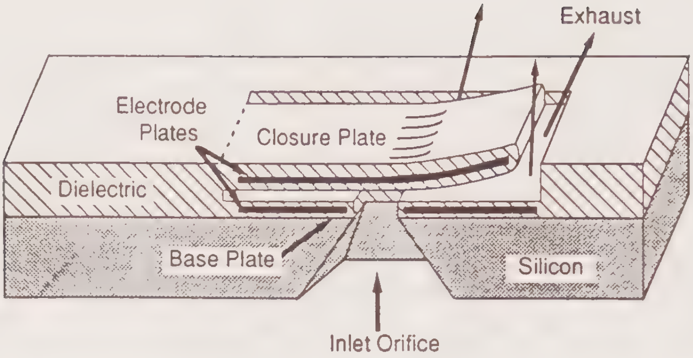

Exhaust
Electrode Plates
Closure Plate
Dielectric
Base Plate
Silicon
Inlet Orifice

FIG. 7.18 Schematic view of a normally open microvalve using electrostatic actuation. A voltage applied across the electrode plates causes the plate to close against the inlet orifice.[72] (©1990 IEEE)

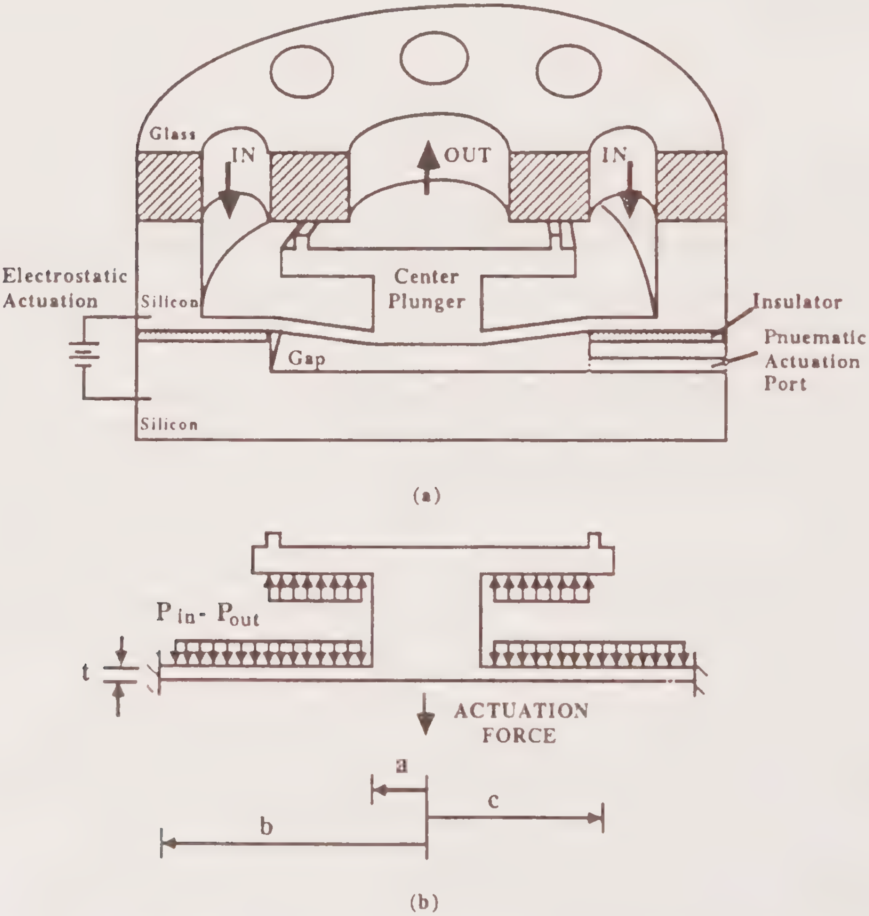

Glass
IN
OUT
IN
Electrostatic Actuation
Silicon
Center Plunger
Insulator
Pnuematic Actuation Port
Gap
Silicon
(a)
$P_{in} - P_{out}$
t
ACTUATION FORCE
a
c
b
(b)

FIG. 7.19 Schematic and cross-sectional views of a pressure-balanced microvalve with electrostatic actuation.[73] (©1993 IEEE)

A microvalve designed to provide an adequate gas flow rate under low pressure has been fabricated.[74] The structure is shown in Fig. 7.20. It is a three-port gas valve consisting of a pair of electrodes covered with an insulation layer and a conductive film of Fe-Ni alloy between them. The middle part of the film is elastically bent by the applied voltage between one of the electrodes and the film. By switching the applied voltage, the two outlets can be opened alternately. A gas flow rate of 10 ml/min was achieved under a pressure difference of 60,000 Pa (9 psi) when the applied voltage was 100 V.

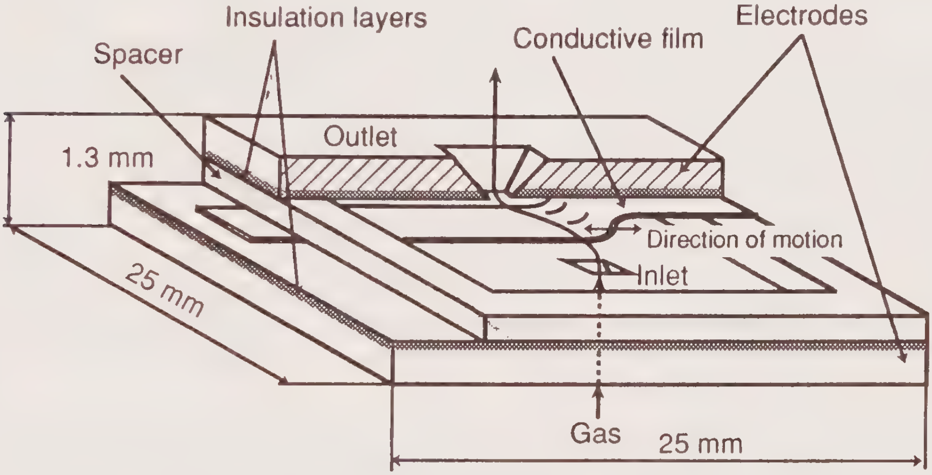


FIG. 7.20 Schematic view of a high-conductance microvalve where a flexible, conductive film is actuated electrostatically to move across an inlet orifice, closing the valve.[74] (©1994 IEEE)

Bimetallic actuators. The structure of a bimetallic actuator is shown in Fig. 7.21. The actuator consists of a center boss, a circular diaphragm made of bimetallic materials, and a circular heater.[75] By varying the electrical power of the heater and thus the temperature of the bimetal, the displacement of the center boss can be controlled. The generated pressure P_{bm} is proportional to the difference between the thermal expansion coefficients γ of the two materials, and the temperature difference ΔT, as

$$P_{bm} \propto (\gamma_a - \gamma_b)\,\Delta T \quad . \tag{7.8}$$

Strong force and reasonable displacement can be expected by using the appropriate combination of the two materials. A bimetallic actuator, either as a beam or diaphragm, tends to apply a concentrated force rather than a distributed actuation pressure. The equivalent generated pressure can be approximated by dividing the force generated by a circular bimetallic diaphragm by the diaphragm area. Although

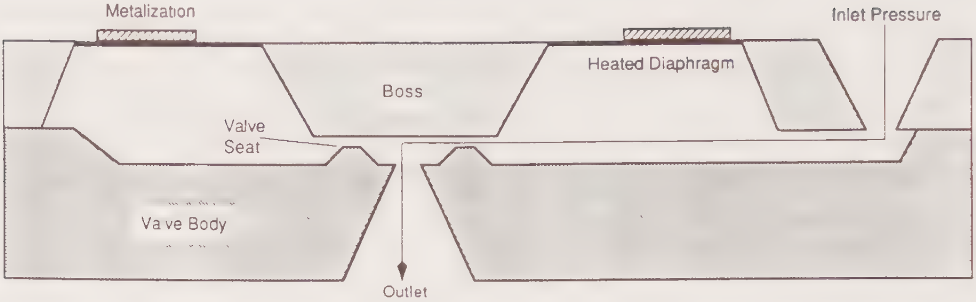


FIG. 7.21 Cross section of a micromachined, bimetallically actuated diaphragm. The inlet pressure is applied on both sides of the actuation diaphragm, lowering the force required to operate the valve.[75]

a number of combinations of materials can be used for the bimetallic actuator, the use of a silicon diaphragm and an aluminum metallic layer is one of the most attractive. For a 3-mm diameter actuator with 10 μm thick aluminum and silicon layers, with simply supported (hinged) outer boundary conditions and a 100°C temperature rise, the equivalent generated pressure is about 34,000 Pa (5 psi). The relaxation time is also the dominant factor in the response time.

Bimetallic beams have been used to provide force and displacement as integrated actuation elements; however, cantilever devices have shown insufficient force for use in most valve applications.[76] A recent device[77] uses a bimetallic cantilever to deflect a stream of liquid, which creates a pressure drop, which in turn opens a diaphragm valve. In effect this multiplies the force available from the cantilever, but makes a valve configuration that may not be generally applicable.

The use of bimetallically driven diaphragms considerably increases the force available compared to a cantilever beam while simultaneously achieving large displacements and symmetric, vertical travel. High-performance diaphragm valves have been developed with these integral actuators to control gas pressure or flow by the direct application of electrical power.[78] In many applications it is useful or necessary to have a normally closed valve which can shut off a flow stream if power is interrupted to the system.

A normally closed valve (see Fig. 7.22) consists of a diaphragm actuator with a central boss that mates to an etched silicon valve body.[79] The actuator is a circular silicon diaphragm with integral diffused resistors that acts as one element of the bimetallic structure. An annular aluminum region on that diaphragm forms the other element. By varying the electrical power dissipated in the resistors and thus the temperature of the silicon diaphragm, the displacement of the central boss can be controlled. The actuator chip and the valve body chip make a very compact valve structure without the large, external actuators common in previously developed valves.

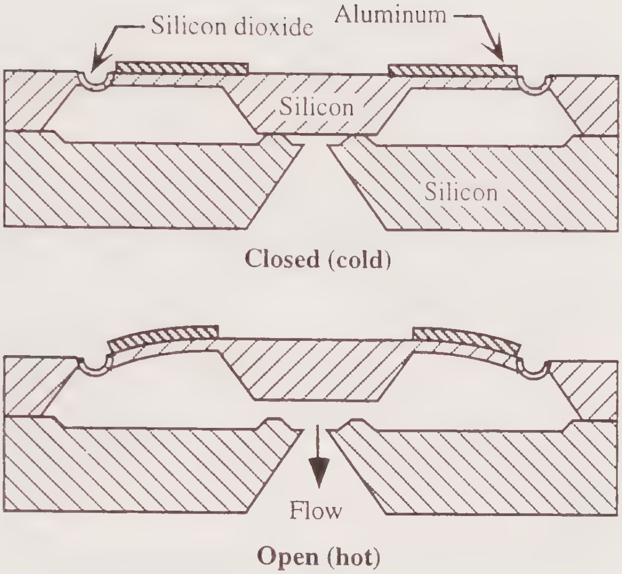


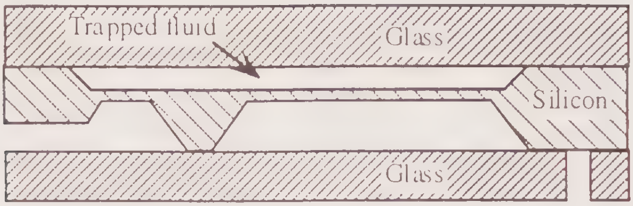
FIG. 7.22 Schematic view of the operation of the bimetallic microvalve. The silicon dioxide hinge provides the correct mechanical boundary conditions for the valve and reduces heat loss to the silicon frame.[79]

Thermopneumatic actuators. A thermopneumatic actuator has a sealed pressure chamber and movable diaphragm. In a typical actuator, the liquid within the chamber is heated by power dissipation in a resistor incorporated in the chamber. The vapor pressure P_{tp} generated in the chamber is given by

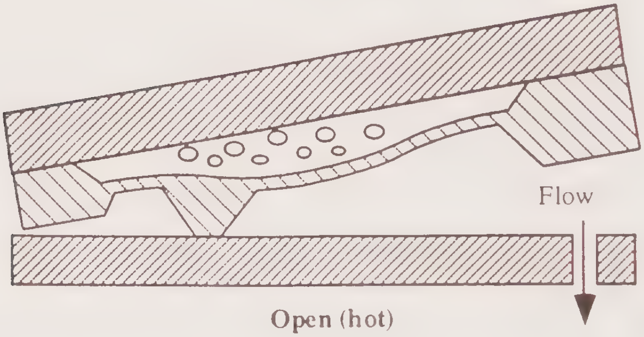
$$P_{tp} = P_0 \exp\left(\frac{-L_0}{RT}\right) , \tag{7.9}$$

where P_0, L_0, R, and T are the initial pressure, the latent heat of vaporization, gas constant, and temperature, respectively. The turn-on time of the actuator depends on the heat capacitance and the available power of the heater, while the relaxation time is determined by the heat transfer to the external world.

A thermopneumatically driven microvalve using a sealed pressure cavity filled with a liquid is shown in Fig. 7.23.[80] A resistive heater was formed on the glass substrate. Using the liquid-gas phase system of methyl chloride, the pressure within the chamber can be increased by 0.7 kgf/cm^2 within 5 ms by applying 200 mW. The dynamic range of this valve can be optimized by adjusting the cavity shape, the boiling point of the control liquid, and the quantity of both gas and liquid in the chamber. Both normally closed and normally open microvalves, which can control liquids at 208 kgf/cm^2 or can control nitrogen gas up to 15 liters/min at 7.1 kgf/cm^2, have been fabricated.[81]



Closed (cold)



Open (hot)

FIG. 7.23 Schematic view of the operation of a microvalve based on the expansion of a heated fluid trapped in a chamber. This valve can be modified to provide either normally open or normally closed operation.[79]

Electromagnetic actuator. An electromagnetically driven microvalve is illustrated in Fig. 7.24.[82] The valve cap is made of an Ni-Fe alloy, a soft magnetic material, supported by a spring. It moves vertically in the magnetic field applied by the external electromagnet. This valve was designed as a flow regulator for a high vacuum application. The valve is driven by a 0.1 Hz to 100 Hz rectangular magnetic field, and the gas flow rate is controlled between 3×10^{-5} torr-liter/min and 2.4×10^{-3} torr-liter/min at a pressure of 4.6×10^{-7} torr. The microvalve can be mounted in a small tube with an external coil mounted outside the tube.

A microvalve activated by a combination of electromagnetic and electrostatic forces has also been fabricated.[83] The structure is illustrated in Fig. 7.25. It consists of a gas flow inlet having a counter electrode, a deflectable membrane coated by a metal conductor, and two permanent magnets (2.5 kG). A current applied to the membrane perpendicular to the magnetic field causes an opening and closing deflection according to the Lorentz force. A voltage can be applied between the membrane and the counter electrode to keep the valve closed electrostatically against a specified gas pressure with low power consumption. A flow of up to 3 ml/min at 0.16 kgf/cm^2 was controlled. Current pulses of 200 mA and voltage of 30 V were typically applied for electromagnetic and electrostatic actuation. The response time was below 0.4 ms.

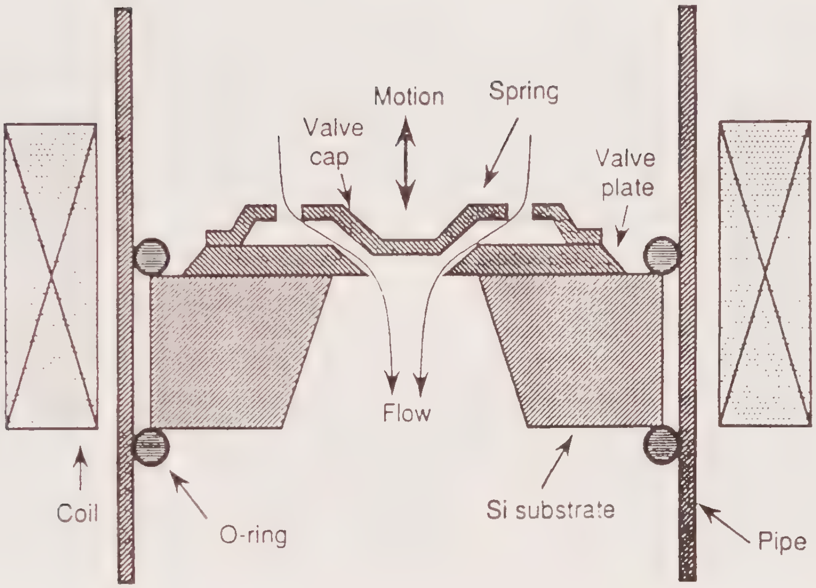

Motion
Spring
Valve cap
Valve plate
Flow
Coil
O-ring
Si substrate
Pipe

FIG. 7.24 Cross-sectional view of a microvalve actuated with magnetic forces. The valve cap is NiFe material and is driven by coils outside the pipe.[82] (©1993 IEEE)

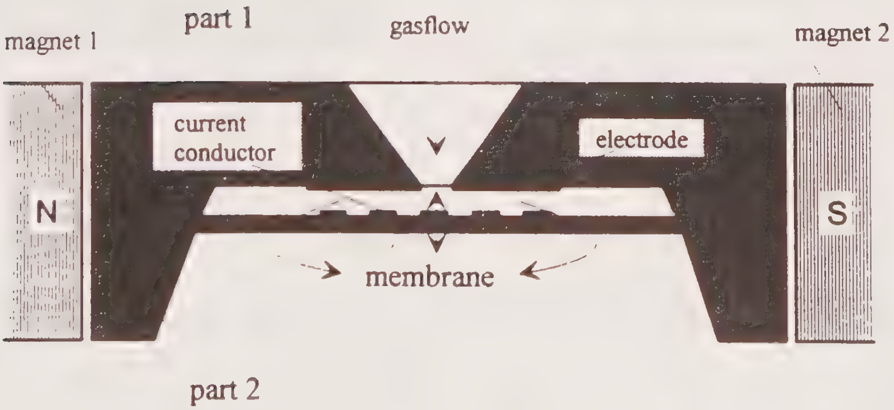

magnet 1
part 1
gasflow
magnet 2
current conductor
electrode
N
S
membrane
part 2

FIG. 7.25 Cross-sectional view of a microvalve with combined electromagnetic and electrostatic actuation.[83] (© Elsevier Science)

7.7.3 Check valves

Check valves are microvalves in which the pressure difference across the valve actuates the valve instead of an independently controllable actuator. Check valves with very small leakage under reverse applied pressure and a large forward-to-reverse flow conductance ratio are required for use in micropumps. The response time, which means transition time during open-to-close or close-to-open, is also an important parameter of the valve. Check valves tend to be small compared to valves with integrated or external actuators.

A wide variety of check valve structures that have been developed have been reported in a review article by Shoji and Esashi.[84] Many valves have been developed for use in integrated micropumps, which will be discussed later. The check valve shown in Fig. 7.26 has a simple cantilever beam structure.[85]

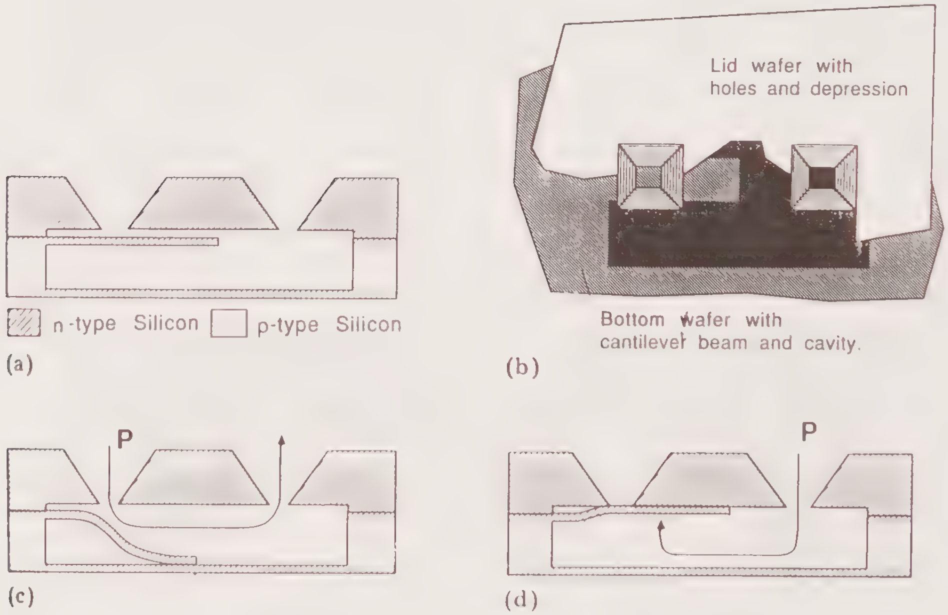

Lid wafer with holes and depression
n-type Silicon
p-type Silicon
(a)
Bottom wafer with cantilever beam and cavity.
(b)
P
(c)
P
(d)

FIG. 7.26 Cross sectional views of a cantilever-type check valve showing: (a) the valve at rest, (b) a schematic top view of the valve, (c) the valve under forward pressure, and (d) the valve under backward pressure.[85] (© Elsevier Science)

Other devices with somewhat similar structure have been fabricated using surface micromachining techniques.[86] In this device, a round polysilicon valve sheet is suspended by four thin polysilicon beams which are attached to the silicon substrate. A bulk micromachined check valve shown in Fig. 7.27 consists of a V-groove with a thin slit, which resembles the valve of a human heart.[87] Other check valves have been made with sheets of polymer such as polyimide or silicon rubber stretched over silicon valve seats.

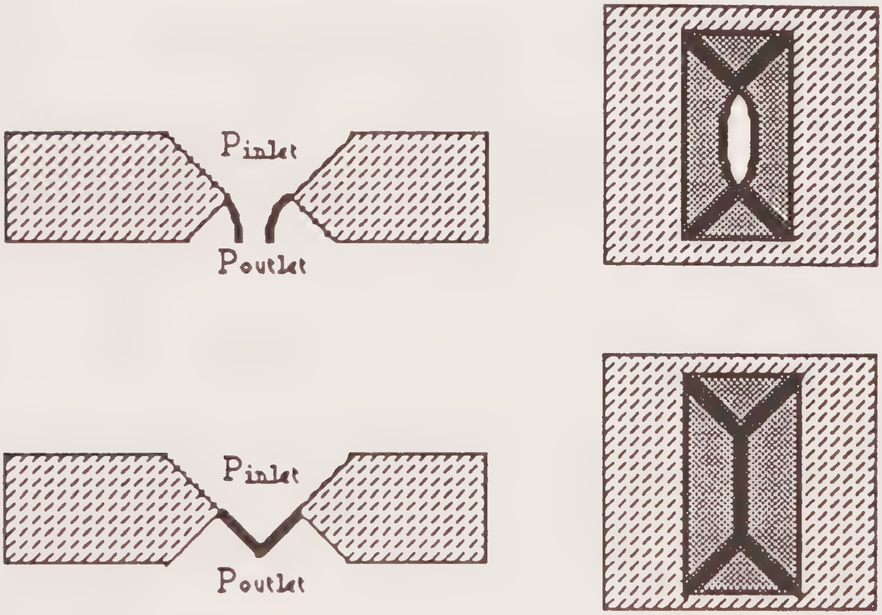


FIG. 7.27 Schematic views of a check valve structure that functions in a manner similar to that in a heart.[87] (©1991 IEEE)

7.7.4 Micropumps

7.7.4.1 Mechanical pumps

Micropumps are classified into two groups: mechanical pumps with moving parts and nonmechanical pumps without moving parts. Two kinds of mechanical micropumps have been developed: reciprocating and peristaltic. Many micropumps of each type using different actuators have been fabricated by micromachining.[88] The check valve and the actuator play very important roles in setting the maximum flow rate and the output pressure of the pump. Many of the same types of actuators used in making individual, active, valves can be used, including piezoelectric, pneumatic, electrostatic, and thermopneumatic. The maximum output pressure of the micropump depends directly on the available force of the actuators used.

In the beginning of the 1980s at the University of Twente, a micropump was created by applying bulk micromachining of silicon for the pump body, and anodic bonding of silicon to glass. The pump consists of three active piezoelectric bimorph valves and works in a peristaltic manner. Van Lintel et al.[89] reported on a pump with passive silicon check valves, and a piezoelectric monomorph glued onto a glass diaphragm as the driver (see Fig. 7.28). In this reciprocating displacement type of pump, the two functions (liquid displacement and valve action) are performed by separate components.

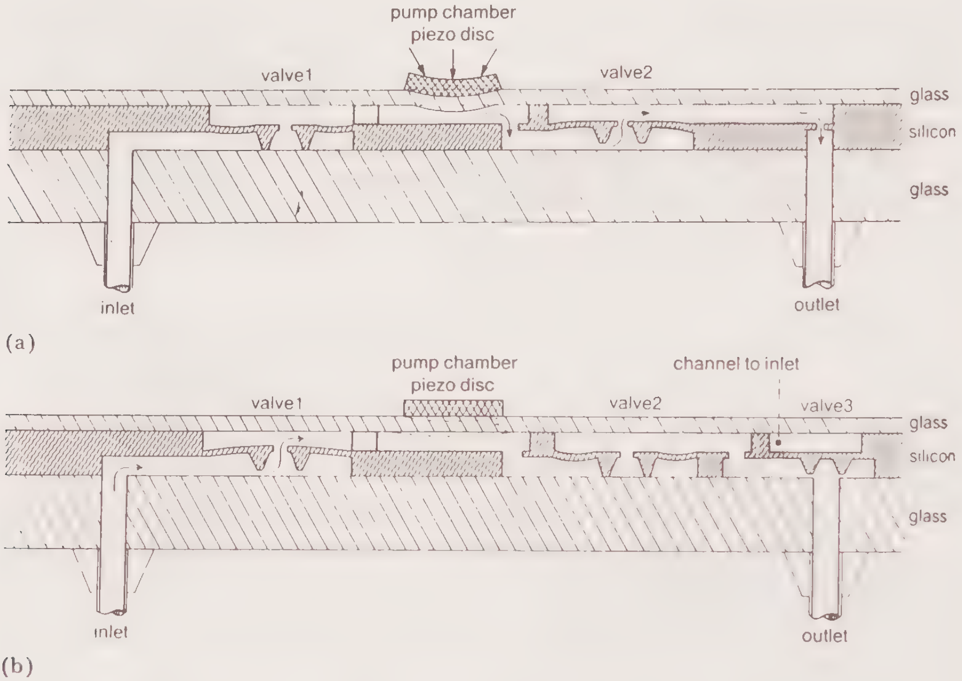

pump chamber
piezo disc
valve1
valve2
glass
silicon
glass
inlet
outlet
(a)
pump chamber
piezo disc
channel to inlet
valve1
valve2
valve3
glass
silicon
glass
inlet
outlet
(b)

FIG. 7.28 Schematic views of two types of piezoelectric pumps: (a) with two valves and (b) with three valves.[89] (© Elsevier Science)

Smits[90] developed a similar type of pump with active, piezoelectrically driven valves glued onto glass diaphragms and functioning in a peristaltic manner. The principle of this type of piezoelectric peristaltic micropump is illustrated in Fig. 7.29.[91] It has three active valves and pumping is carried out when these valves are driven by controlled signals to achieve peristaltic motion. The pumping rate was reported to be 100 μl/min when the driving voltage was 80 V. A negative voltage was used to bend the valves down during their closed period. The maximum pressure was 0.6 mH_2O. The flow was proportional to a driving frequency up to 15 Hz.

In the middle 1980s at Tohoku University, Japan, Esashi and Shoji started research on microsystems for chemical analysis. They created pumps with polysilicon valves

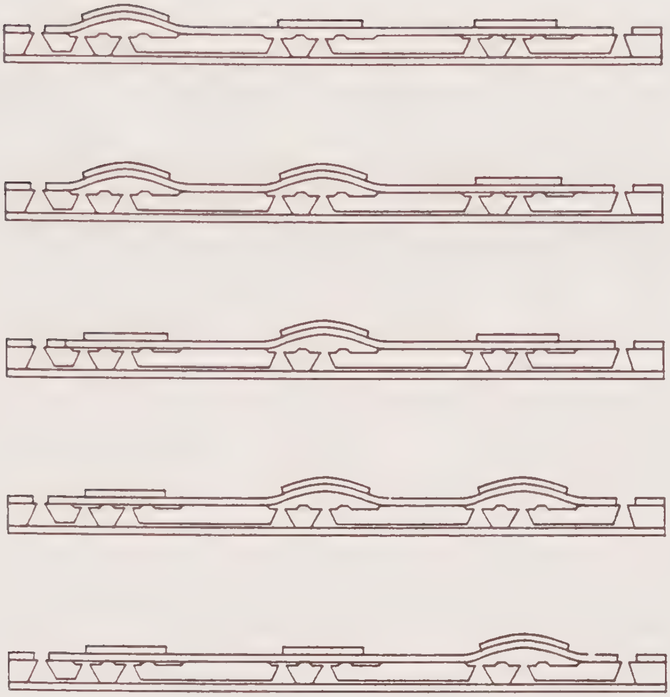


FIG. 7.29 Schematic views showing the operation of a piezoelectric peristaltic micropump.[90] (© Elsevier Science)

fabricated by surface micromachining. These pumps are driven by means of a stack of piezoelectric disks.[92]

Piezoelectric micropumps can be made without using check valves, but with nozzles or restrictors that act in a similar way. One disadvantage of a micropump using check valves is the restriction on the driving frequency due to slow actuation of the check valves. This type of micropump can be driven at frequencies of hundreds of hertz. Liquid flow rates of about 3 ml/min and pump pressures of about 1.0 mH_2O have been obtained with a driving frequency of 300 Hz and applied voltage of 140 V. The maximum pressure of this pump was 2.5 mH_2O. Another remarkable advantage was the capability of pumping gas. Pumping of air was demonstrated under 6 kHz driving frequency and 20 V voltage to provide a flow of 35 ml/min. This device can also pump a fluid with solid particles, which isn't possible using check valves.

A pneumatically driven micropump was fabricated using the LIGA process.[93] The micropump is made with polyimide diaphragms as shown in Fig. 7.30. The check valve consists of a titanium wire structure and a polyimide membrane. A maximum flow rate of 80 μl/min and a maximum pressure of 0.47 mH_2O were obtained for air pumping at a driving frequency of 5 Hz. The use of a pneumatic actuator severely limits the frequency response of this pump.

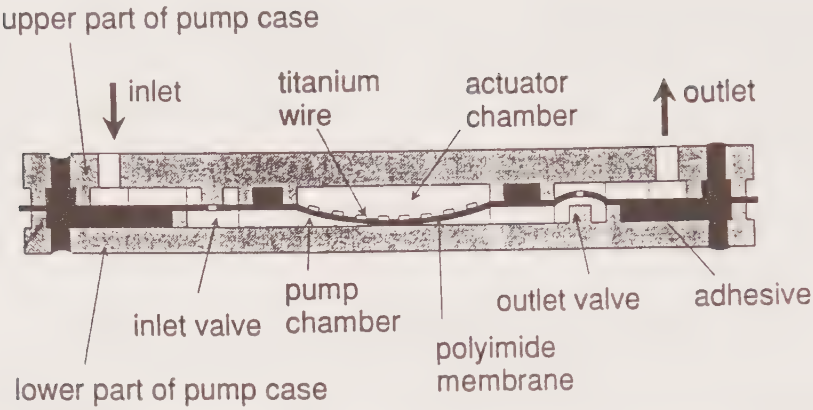

upper part of pump case
inlet
titanium wire
actuator chamber
outlet
pump chamber
inlet valve
outlet valve
adhesive
polyimide membrane
lower part of pump case

FIG. 7.30 Schematic view of a micropump fabricated with molded plastic parts and a polyimide membrane with a titanium wire heater.[93] (© Elsevier Science)

van de Pol et al.[94] created a reciprocating displacement type of micropump that is thermopneumatically driven. A small amount of air contained in a cavity is raised in temperature by resistive heating. The induced pressure is used to actuate a diaphragm, which in turn displaces the liquid. Bulk-type check valves are used. The pumping rate was in the range of 0 to 50 μl/min at frequency of about 5 Hz. Similar results were obtained using laser light to heat up the chambers. A thermopneumatically driven peristaltic micropump having three active pressure chambers with flexible membranes has also been made.[95] In this case, a microresistive heater was used to change the temperature of the pressure chamber. A similar thermopneumatic micropump was fabricated using a resistive microheater, as shown in Fig. 7.31.[96]

An electrostatically driven micropump has been made which consists of four stacked silicon wafers. The check valves are fabricated using cantilevers formed on

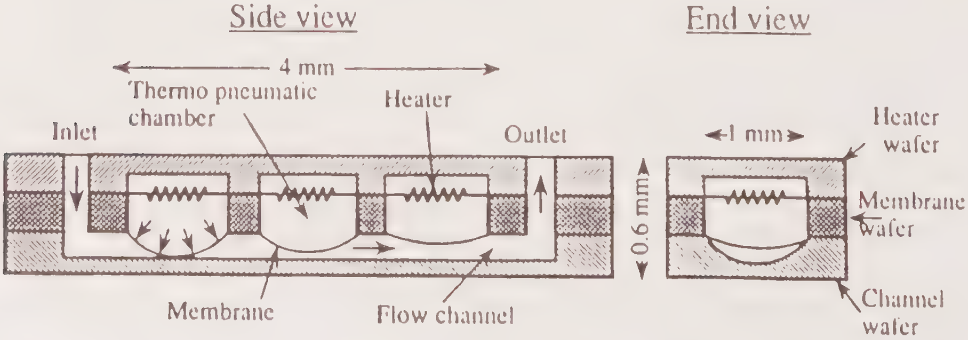

Side view
End view
4 mm
Thermo pneumatic chamber
Heater
Inlet
Outlet
Membrane
Flow channel
1 mm
0.6 mm
Heater wafer
Membrane wafer
Channel wafer

FIG. 7.31 Cross-sectional view of a thermopneumatic peristaltic micropump.[95] (©1992 IEEE)

the two lower wafers. This can be fully fabricated in a batch process. The driving frequency range is from 1 to 1000 Hz. At a frequency of 400 Hz, a flow rate of 350 μl/min and a maximum pressure of 2.4 mH_2O were obtained.

7.7.4.2 Nonmechanical pumps

It is possible to move liquids and constituents in liquids without using any moving, mechanical parts. These techniques use a variety of electric-fluid interaction phenomenon to generate forces on the fluids. Electrohydrodynamic effects, electro-osmotic phenomena, and ultrasonic effects can be used in nonmechanical micropumps, which do not have moving parts and are very attractive in a variety of applications.

Fluidic motion can be induced by traveling flexural waves for the transport of liquids (Fig. 7.32). This is the so-called ultrasonically driven micropump.[97] A liquid moves in the direction of wave propagation with a speed proportional to the square of the acoustic amplitude. A maximum speed of 130 μm/s was observed with a drive voltage of 7.1 V rms at 3.5 MHz.

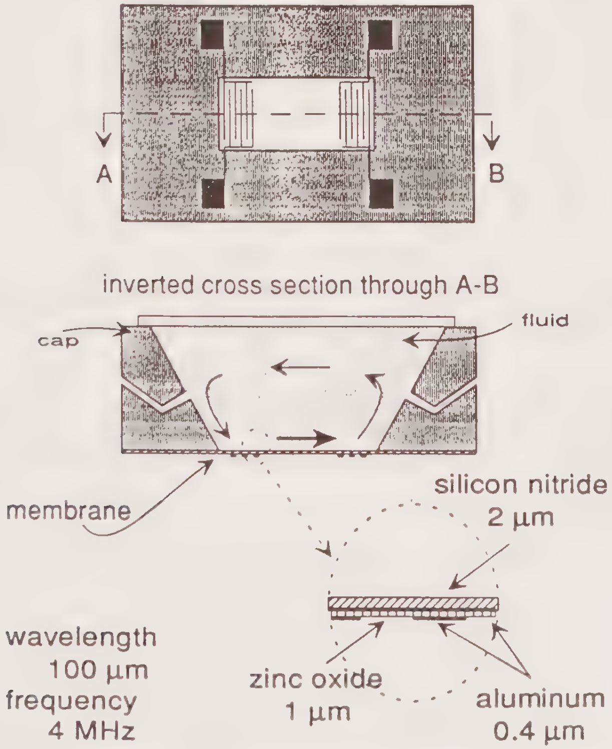

A
B
inverted cross section through A-B
cap
fluid
membrane
silicon nitride
2 µm
wavelength
100 µm
frequency
4 MHz
zinc oxide
1 µm
aluminum
0.4 µm

FIG. 7.32 Schematic view of a lamb-wave membrane pump. A cap confines the fluid in a 250-m-deep well.[97] (©1991 IEEE)

By micromachining a manifold of flow channels in a planar substrate, it is possible to fabricate a system capable of sample injection, pretreatment, and separation for an analysis system without using conventional pumps. The linear flow rates obtained are well suited for use in chemical analysis systems, since the flow rate of solvent is approximately independent of capillary dimensions. A typical injection system for an analysis system consists of inlets, or reservoirs, at the heads of interconnected capillary channels. The application of a voltage between any two inlets causes electro-osmotic pumping of fluid along those channel segments between the inlets. Valveless switching of fluid flow between channels can be achieved by switching the voltages applied to each channel.

Electrophoresis. The basis for the electrophoretic separation process is the differential migration of sample ions relative to solvent molecules under the influence of an externally applied electric field. Whenever an electric field is applied to electrolyte solutions, ions in the solution will be accelerated until the electric force is balanced by a frictional force on a picosecond time scale.

Electro-osmosis. In contrast to electrophoresis, electro-osmosis is a macroscopic phenomenon and involves the pumping of fluid through a capillary tube by an externally applied electric field.[98] A prerequisite for electro-osmosis is the presence of immobilized surface charges at the capillary wall in contact with an electrolyte solution. The surface charge density might be due either to ionizable groups (e.g., deprotonation of silanol groups in the case of glass or fused silica capillaries) or caused by strong adsorption of charged species which are present in the buffer. This surface charge leads to the formation of an electric double layer by attracting oppositely charged ions from the buffer and therefore leads to concentration and charge density gradients in the immediate vicinity of the wall. A manifold of capillary channels can be fabricated in a planar glass substrate. Since the solvent flow can be controlled along a specified capillary by applying appropriate voltages, valveless switching of liquid flow between capillaries can be obtained.

If an electrical field is now applied in the axial direction, a volume force will be exerted on the fluid. However, the net charge density in the buffer is significantly different from zero only in a thin annular region close to the wall; therefore only this region will experience the external field and start moving according to the field direction and sign of the accumulated charge. Due to the radial velocity gradient formed this way, the adjacent fluid annuli will be accelerated by the momentum transfer caused by viscous forces until the velocity gradient approaches zero across the whole radius of the capillary. Numerical simulations have shown that this process takes place on a time scale between 100 μs and 1 ms.[99] After that time, the whole fluid inside the capillary moves at a constant speed, i.e., the resulting flow profile across the capillary is of rectangular shape.[100]

Electrohydrodynamic micropumps. With electrohydrodynamic actuation, the pumping of a fluid is limited to liquids of low conductivity and dielectric liquids such as organic solutions. A first attempt at constructing an electrohydrodynamic micropump for the displacement of charged liquids was described by Bart et al.[101] Electrohydrodynamic motion in a dielectric liquid is a complex process. The motive force of the pumping is the coulomb force exerted on the charges between the two pump electrodes. Under high field conditions, the injection of ions from electrodes to the liquid occurs. This phenomenon can be controlled by the electrochemical reactions at the electrode–liquid interface and is dependent on the composition and the geometry of the electrodes. Meshlike parallel electrodes were utilized for this purpose in one version of such a pump.[102] Since the injected ions are moved by the attractive electrostatic force, the gap between the electrodes should be as small as possible. The maximum flow rate was 12 ml/min and the maximum output pressure was 0.24 mH_2O at 700 V. In this case, the conductivity had to be in the range between 10^{-4} S/cm and 10^{-9} S/cm.

7.7.5 Conclusions

A substantial portion of the development of actuators for micromachined devices has focused on valves and pumps for miniature systems because there is a perceived need and market for these devices. The relatively low output force or pressure from the miniature elements has thus far limited the commercial impact of these devices. For liquid analysis systems, nonmechanical pumping designs using electro-osmosis or electrophoresis show great promise because the performance of these techniques is suitable for miniature systems.

References

1. S. Middelhoek and A. C. Hoogerwerf, "Smart sensors: when and where?" *Sensors Actuators*, **8**, 39–48 (1985).
2. S. Middlehoek and D. J. W. Noorlag, "Three-dimensional representation of input and output transducers," *Sensors Actuators*, **2**, 29–41 (1981/82).
3. K. E. Petersen, "Silicon as a mechanical material," Proc. IEEE, **70** (5), 420–457 (May 1982).
4. C. Bourgeois, J. Hermann, N. Blanc, N.F. de Rooij, and F. Rudolf, "Determination of the elastic temperature coefficients of monocrystalline silicon," *Tech. Dig. Transducers '95*, Vol. 2, 92–95 (1995).
5. K. G. Lyon, G. L. Salinger, C. A. Swenson, and J. K. White, *J. Appl. Phys.*, **48**, 865 (1977).
6. C. S. Smith, "Piezoresistance effect in germanium and silicon," *Phys. Rev.*, **94**, 42–49 (Apr. 1954).
7. O. N. Tufte, P. W. Chapman, and D. Long, "Silicon diffused element piezoresistive diaphragms," *Appl Phys*, **33**, 3322–3327 (Nov. 1962).

8. Y. Kanada, "Piezoresistance effect in silicon," *Sensors Actuators*, **28**, 83–91 (1991).
9. C. Herring, "Transport properties of a many-valley semiconductor," *Bell System Tech J.*, **34**, 237–296 (Mar. 1955).
10. Y. Kanada, "A graphical representation of the piezoresistance coefficients in silicon," *IEEE Trans Elec. Dev.*, **ED-29**, 64–70 (Jan. 1982).
11. W. P. Mason and R. N. Thurston, "Use of piezoresistive materials in the measurement of displacement, force, and torque," *J. Acoust. Soc. Amer.*, **29**, 1096–1101 (Oct. 1987).
12. C. S. Sander, J. W. Knutri, and J. D. Meindl, "A monolithic capacitive pressure sensor with pulsed-period output," *IEEE Trans. Electron. Dev.*, **ED-27**, 927–930 (May 1980).
13. Y. S. Lee and K. D. Wise, "A batch-fabricated silicon capacitive pressure transducer with low temperature sensitivity," *IEEE Trans. Electron. Dev.*, **ED-29**, 42 (Jan. 1982).
14. H. L. Chau and K. D. Wise, "Scaling limits in batch-fabricated silicon pressure sensors," in *Tech. Dig.Transducers '85* , 174–177 (June 1985).
15. J. S. Smith, L. Bowman, and J. P. Meindl, "Analysis, design and performance of a capacitive pressure sensor IC," *IEEE Trans. Biomed. Eng.* **BME-33**, 174–177 (1986).
16. Samaun, K. D. Wise, and J. B. Angell, "An IC piezoresistive pressure sensor for biomedical instrumentation," *IEEE Trans. Biomed. Eng.* **BME-20**, 101–l09 (1973).
17. W. H. Ko, M. H. Bao and Y. D. Yang, "A high sensitivity integrated circuit capacitive pressure transducer," *IEEE Trans. Electron. Dev.* **ED-29**, 48–56 (1982).
18. M. E. Behr, C. F. Bauer and J. M. Giachino, "Miniature silicon capacitance absolute pressure sensor," *J. Mech. Eng.* **C196/8l**, 255–260 (1981).
19. Y. Zhang, S. B. Crary , and K. D. Wise, "Pressure sensor design and simulation using the CAEMEMS-D module," in *Tech. Dig. IEEE Solid-State Sensors and Actuator Workshop,* 32–35 (1990).
20. Y. Zhang and K. D. Wise, "An ultra-sensitive capacitive pressure sensor with a bossed dielectric diaphragm," in *Tech. Dig. IEEE Solid-State Sensor and Actuator Workshop,* 205–208 (1994).
21. J. J. Sniegowski, "Design and fabrication of the polysilicon resonating beam force transducer," PhD Dissertation, University of Wisconsin-Madison (1991).
22. K. I. Ikeda, H. Kumayama, T. Koboyashi, T. Wanabe, T. Nishikawa, T. Yoshida, and K. Harada, "Silicon pressure sensor integrates resonant strain gauge on diaphragm," *Sensors Actuators* **A21-23** 144–150 (1990).
23. H. Guckel, C. Rypstat, M. Nesnidal, J. D. Zook, D. W. Burns, and D. K. Arch, "Polysilicon resonant microbeam technology for high-performance

sensor applications," in *IEEE Solid-State Sensor and Actuator Proc.*, 153–156 (1992).

24. L. M. Roylance and J. B. Angell, "A batch-fabricated silicon accelerometer," *IEEE Trans. Electron. Dev.*, **ED-26**, 1911–1917 (1979).
25. Henry V. Allen et al., "Accelerometer systems with self-testable features," *Sensors Actuators,* **20** 158–161 (1989).
26. S. C. Terry, "A miniature silicon accelerometer with built-in damping," *IEEE Solid-State Sensor and Actuator Workshop*, 114–116 (1988).
27. T. Tschan, N. de Rooij and A. Bezinge, "Oil damped piezoresistive silicon accelerometers," *in Proc. 6th Int. Conf. Solid State Sensors and Actuators (Transducers '91),* 112–114 (1991).
28. F. Rudolf et al., "Precision accelerometers with mg resolution," *Sensors Actuators*, **A2l-A23**, 297–302 (1990).
29. S. Suzuki et al., "Semiconductor capacitance-type accelerometer with electrostatic servo technique," *Sensors Actuators*, **A2l-A23**, 316–319 (1990).
30. W. Henrion et al., "Wide dynamic range direct digital accelerometer," in *IEEE Solid-State Sensor and Actuator Workshop*, 153–157 (1990).
31. S. C. Chang et al., "Resonant-bridge two-axis accelerometer," *Sensors Actuators*, **A2l-A23**, 342–345 (1990).
32. B. Boxenhorn and P. Greiff, "Monolithic silicon accelerometer," *Sensors Actuators*, **A21-A23**, 273–277 (1990).
33. F. Rudolph, "Precision accelerometers with μg resolution," *Sensors Actuators*, **A21-A23**, 297–301 (1990).
34. W. Henrion et al., "Wide dynamic range direct digital accelerometer," *IEEE Solid-State Sensor and Actuator Workshop*, 153–157 (1990).
35. L. Ristic et al., "Surface micromachined polysilicon accelerometer," in *IEEE Solid-State Sensor and Actuator Workshop*, 118–121 (June 1992).
36. R. S. Payne and K.H. Dinsmore, "Surface micromachined accelerometer: A technology update," *Proc. SAE 1991*, 127–135 (1991).
37. K. Chau et. al., "Integrated force balanced capacitive accelerometers," in *IEEE Solid State Sensor and Actuator Workshop*, 593–596 (1995).
38. S. C. Chang et al., "Resonant-bridge two-axis accelerometer," *Sensors Actuators*, **A2l-A23**, 342–345 (1989).
39. S. B. Waltman and W. J. Kaiser, "An electron tunneling sensor," *Sensors Actuators*, **19**, 201–210 (1989).
40. T. W. Kenny, S.B. Waltman, J.K. Reynolds and W.J. Kaiser, "A micromachined silicon electron tunneling sensor," in *Proc. IEEE Micro Electro Mechanical Systems Conf.*, 192–195 (1994).
41. A. A. Baski, T. R. Albrecht and C. F. Quate, "Tunneling accelerometer," *J. Microscopy* **157**, 73–76 (1988).

42. T. W. Kenny, S. B. Waltman, J. K. Reynolds and W. J. Kaiser, "A micromachined silicon electron tunneling sensor," in *Proc. IEEE Micro Electro Mechanical Systems Conf.*, 192–195 (1994).
43. T. W. Kenny, S. B. Waltman, J. K. Reynolds and W. J. Kaiser, "Micromachined silicon tunnel sensor for motion detection," *Appl. Phys. Lett.* **58** 100–102 (1991).
44. T. W. Kenny, W. J. Kaiser, J.K. Reynolds, J. A. Podosek, H. K. Rockstad, E. C. Vote, and S. B. Waltman, "Electron tunnel sensors," *J. Vac. Sci Tech*, **A10**, 2114–2118 (1992).
45. H. K. Rockstad, T. W. Kenney, J. K. Reynolds, W. J. Kaiser, and T. B. Gabrielson, "A miniature high-sensitivity broad-band accelerometer based on electron tunneling transducers," in *Proc. Transducers '93*, 836–839 (1993).
46. J. .J. Yao, S. C. Amey, and N. C. MacDonald, "Fabrication of high-frequency, two-dimensional nanoactuators for scanned probe devices," *J. MicroElectroMechanical Syst.*, **1**, 14–22 (1992).
47. D. Kobayashi, T. Hirano, T. Furuhata, and H. Fujita, "An integrated lateral tunneling unit," in *Proc. Micro Electro Mechanical Systems*, 214–219 (1992).
48. K. S. Udell et al., "Microsensors for heat transfer and fluid flow measurements," *Exp. Therm. Fluid Science*, **3**, 52–59 (1990).
49. A. W. van Herwaarden and P. M. Sarro, "Floating-membrane thermal vacuum sensor," in *Tech. Digest, Transducers '87*, 287–290 (1987).
50. B. W. van Oudheusen, "Silicon thermal flow sensors," *Sensors Actuators* **A31**, 9–16 (1992).
51. M. Esashi, "Microflow sensor and integrated micro oxygen sensor using it," in *Tech.Dig. Transducers '91* 34–37 (1991).
52. D. Moser, R. Lenggenhager, and H. Baltes, "Silicon gas flow sensors using industrial CMOS and bipolar IC technology,"*Sensors Actuators* **A27**, 577 (1991).
53. M. A. Schmidt, R. T. Howe, S. D. Senturia, and J. H. Haritonidas, "Design and calibration of a microfabricated floating-element shear stress sensor," *IEEE Trans. Elec. Dev.* **ED-35**(6), 750–756 (1988).
54. D. D. Denton, S. D. Senturia, E. S. Anolick, and D. Scheider, "Fundamental issues in the design of polymeric capacitive moisture sensors," in *Tech. Dig. Third Int. Conf. on Solid-State Sensors and Actuators*, 202–205 (1985).
55. T. S. J. Lammerink, M. Elwenspoek, and J. H. J. Fluitman, "Integrated micro-liquid dosing system," in *Proc. IEEE-MEMS Workshop* 254–259 (1993).
56. C. Yang, M. Kimmel, and H. Soeberg, "A transit-time flow meter for measuring milliliter per minute liquid flow," *Rev. Sci. Instr.* **59**(2) 314–317 (1988).

57. C. Yang and H. Soeberg, "Monolithic flow sensor for measuring millilitre per minute liquid flow," *Sensors Actuators*, **A33**, 143–153 (1992).
58. V. Gass, B. H. van der Schoot, and N. F. de Rooij, "Nanofluid handling by micro-flow-sensor based on drag force measurements" in *Proc. IEEE-MEMS Workshop,* 167–172 (1993).
59. T. Nishimoto, S. Shoji, and M. Esashi, "Buried piezoresistive sensors by means of MeV ion implantation," in *Tech. Dig. Transducers '93,* 796–799 (1993).
60. J. T. Suminto, G.-J. Yeh, T. M. Spear, and W. H. Ko, "Silicon diaphragm capacitive sensor for pressure, flow, acceleration and altitude measurements," *Proc. Transducers '87 Conference,* 336–339 (1987).
61. R. J. G. Carr, "Fibre optic sensors for the characterization of particle size and flow velocity," *Sensors Actuators* **A21-A23** 1111–1117 (1990).
62. S. C. Terry, "A Gas Chromatography System Fabricated on a Silicon Wafer Using Integrated Circuit Technology," Stanford Elec. Lab. Technical Report #4603-1, p. 41 (1975).
63. S. C. Terry, J. H. Jerman, and J. B. Angell, "A gas chromatographic air analyzer fabricated on a silicon wafer," *IEEE Trans Electron. Dev.* **ED-26**(12), 1880–1886 (1979).
64. S. Nakagawa, S. Shoji, and M. Esashi, "Integrated fluid control systems on a silicon wafer," in *Micro System Technologies '90*, 793–798 (1990).
65. W. G. Wolber, "Automotive control actuators: an overview," *Automotive Eng.* **92**(8) 55 (1984).
66. M. J. Zdeblick and J. B. Angell, "A microminiature electric-to-fluidic valve," in *Proc. 4th Int. Conf. on Solid-State Transducers and Actuators,* 827–829 (1987).
67. S. C. Terry, J. H. Jerman, and J. B. Angell, "A gas chromatographic air analyzer fabricated on a silicon wafer," *IEEE Trans. Electron Dev.* **ED-26**, 1880–1886 (1979).
68. M. Esashi, "Integrated microflow control systems, *Sensors Actuators* **A21-A23** 161–167 (1990).
69. S. Nakagawa, S. Shoji, and M. Esashi, "A microchemical analyzing system integrated on a silicon wafer," in *Proc. IEEE-MEMS Workshop,* 89–94 (1990).
70. S.C. Terry, A Gas Chromatography System Fabricated on a Silicon Wafer Using Integrated Circuit Technology, Stanford Elec. Lab. Technical Report #4603-1, p. 41 (1975).
71. S. Shoji, M. Esashi, and M. Matsuo, "Prototype miniature blood gas analyzer fabricated on a silicon wafer," *Sensors Actuators*, **14**, 101–107 (1988).
72. T. Ohnstein, T. Fukiura, J. Ridley, and U. Bonne, "Micromachined silicon microvalve," in *Proc. IEEE-MEMS Workshop*, 95–98 (1990).

73. M. A. Huff, J. R. Gilbert, and M. A. Schmidt, "Flow characteristics of a pressure-balanced microvalve," in *Tech. Dig. of Transducers '93,* 98–101 (1993).
74. M. Shikida and K. Sato, "Characteristics of an electrostatically-driven gas valve under high-pressure conditions," in *Proc. IEEE-MEMS Workshop,* 235–240 (1994).
75. H. Jerman, "Electrically-activated micromachined diaphragm valves," in *Tech. Dig. IEEE Sensors and Actuators Workshop,* 65–69 (1990).
76. W. Reithmuller and W. Benecke, "Thermally excited silicon microactuators," *IEEE Trans. Electron. Dev.*, **35** (6), (1988).
77. H.-P. Trah, H. Baumann, C. Doring, H. Goebel, T. Grauer, and M. Mettner, "Micromachined valve with hydraulically actuated membrane subsequent to a thermoelectrically controlled bimorph cantilever," *Sensors Actuators, A*, **39**, 169–176 (1993).
78. J. H. Jerman, Semiconductor Microactuator, U.S. Patent #5,069,419, issued December 1991.
79. P. W. Barth, "Silcion microvalves for gas flow control," in *Tech Dig. Transducers '95*, Vol. 2, 276–279 (1995).
80. M. J. Zdeblick and J. B. Angell, "A microminiature electric-to-fluidic valve," *Tech. Dig. Transducers '87*, 827–829 (1987).
81. M. J. Zdeblick, R. Anderson, J. Jankowski, B. Kline-Schoder, L. Christel, R. Miles, and W. Weber, "Thermopneumatically actuated microvalves and integrated electro-fluidic circuits," in *Tech. Dig. IEEE Solid-State Sensor and Actuator Workshop*, 251–255 (1994).
82. K. Yanagisawa, H. Kuwano, and A. Tapo, "An electromagnetically driven microvalve," in *Tech. Dig. of Transducers '93* 102–105 (1993).
83. D. Bosch, B. Heimhofer, G. Muck, H. Seidel, U.Thumser, and W. Welser, "A silicon microvalve with combined electromagnetic/electrostatic actuation," *Sensors Actuators* **A37-A38**, 684–692 (1992).
84. S. Shoji and M. Esashi, "Microflow devices and systems," *J. Micromech. Microeng.*, **4**, 157–171 (1994).
85. J. Tiren, L. Tenerz, and B. Hok, "A batch-fabricated non-reverse valve with cantilever beam manufactured by micromachinined silicon," *Sensors Actuators* **18**, 389–396 (1989).
86. M. Esashi, S. Shoji, and A. Nakano, "Normally closed microvalve and micropump fabricated on a silicon wafer," *Sensors Actuators* **20**, 163–167 (1989).
87. L. Smith and B. Hok, "A silicon self-aligned non-reverse valve," in *Tech. Dig. Transducers '91*, 1049–1051 (1991).
88. R. Zennerle, M. Richter, F. Brosinger, A. Richter, and H. Sandmaier, "Performance simulation of microminiaturized membrane pumps," in *Tech. Dig. Transducers '93,* 106–109 (1993).

89. H. T. G. van Lintel, F. C. M. van de Pol and S. Bouwstra, “A piezoelectric micropump based on micromachining of silicon,” *Sensors Actuators* **15**, 153–167 (1988).
90. J. G. Smits, “Piezoelectric micropump with three valves working peristaltically,” *Sensors Actuators* **A2l-A23**, 203–206 (1990).
91. See Ref. 90.
92. S. Shoji and M. Esashi, “Micropump for integrated chemical analyzing systems,” in *Tech. Dig. of the 7th Sensor Symposium*, 217–220 (1988).
93. B. Büstgens, W. Bacher, W. Manz, W. K. Schomburg, “Micropump manufactured by thermoplastic molding,” *Proc. Micro Electro Mech. Sys. '94*, 18–21 (1994)
94. F. C. M. van de Pol, H. T. G. van Lintel, M. Elwenspoek and J. H. J. Fluitman, “A thermopneumatic micropump based on micro-engineering techniques,” *Sensors Actuators* **A21-A23**, 198–202 (1990).
95. J. A. Folta, N. F. Raley, and E. W. Hect, “Design, fabrication and testing of a miniature peristaltic membrane pump,” in *Tech.Dig. IEEE Sensors and Actuators Workshop* 185–189 (1992).
96. T. S. J. Lammerink, M. Elwenspoek, and J. H. J. Fluitman, “Integrated micro-liquid dosing system,” in *Proc. IEEE-MEMS Workshop*, 25–29 (1993).
97. R. M. Moroney, R. M. Whilt, and R. T. Howe, “Ultrasonic induced microtransport,” in *Proc. 1991 IEEE MEMS Workshop*, 277–282 (1991).
98. D. J. Harrison, K. Seiler, A. Manz, and Z. Fan, “Chemical analysis and electrophoresis systems integrated on glass and silicon chips,” in *Tech. Dig. IEEE Solid-State Sensor and Actuator Workshop* 110–113 (1992).
99. E. V. Dose and G. Guiochon. *J. Chromatogr.* **65**(2), 263–275 (1993).
100. C. L. Rice and R. Whitehead, *J. Phys. Chem.* **69**, 4017–4024 (1993).
101. S. F. Bart, L. S. Tavrow, M. Mehregany, and- J. H. Larig, “Microfabricated electrohydrodynamic pumps,” *Sensors Actuators* **A2l-A23**, 193–197 (1990).
102. A. Richter and H. Sandmaier, “Electrohydrodynamic pumping and flow measurement,” in *Proc. IEEE-MEMS Workshop*, 99–104 (1991).
103. J. Bryzek, K. Petersen, and W. McCulley, "Micromachines on the march," *IEEE Spectrum,* **31**, 20–31 (1994).
104. P. C. Tack and H. H. Busta, "Stress modeling of micro-diaphragm pressure sensors," in *Tech. Dig. IEEE Sensors and Actuators Workshop* (1986).

CHAPTER 8

Micro-Optical Devices

Hiroyuki Fujita
Hiroshi Toshiyoshi
University of Tokyo

CONTENTS

8.1 Concept of Optical MEMS / 436

8.2 Recent examples of optical MEMS / 439
  8.2.1 Applications of MEMS Fabry-Perot interferometers / 439
  8.2.2 Micromechanical optical switches / 455
  8.2.3 Optical scanners and choppers / 475
  8.2.4 Optical elements using three-dimensional polysilicon structures / 482
  8.2.5 Microlenses and focusing mirrors / 487
  8.2.6 Optical elements made by the LIGA process / 490
  8.2.7 Display / 491
  8.2.8 Micro-optical disk / 495
  8.2.9 Near-field scanning optical microscope / 497
  8.2.10 Optical microencoder and displacement sensor / 497
  8.2.11 Waveguide sensors integrated with MEMS structures / 499

Conclusion / 504

References / 505

8.1 Concept of optical MEMS

Microelectromechanical systems (MEMS) based on integrated circuit (IC) compatible micromachining [1, 2, 3] have advantages denoted by three "M's." Miniaturization is the most popular but multiplicity, which means the batch fabrication capability of many complicated elements, and microelectronics to control motion or to add different functions such as the optical function, are equally important. This chapter deals with the application of micromachine technologies to micro-optical devices. A basic concept making the best use of the advantages is proposed. Recent examples of optical microelectromechanical systems are reviewed.

The invention of lasers opened a wide variety of applications of optical systems. Integrated optics with semiconductor lasers, optical waveguides, mirrors and lenses were quickly developed . Solid-state optical devices based on nonlinear optical properties are capable of deflecting and modulating optical beams without moving parts. Optomechanical devices are advantageous, however, because they have large deflecting or scanning angles and good contrast between on and off states. Microactuators can be utilized in such devices since only small forces are required to move mirrors and shutters. One can obtain optical MEMS by combining microactuators and structures with integrated optics. These are possibilities for applying optical MEMS to switches in optical communication networks, scanners in optical sensors, and focusing in CD and optical disk storage.

Some examples of optical MEMS date back to the 1970s. In 1977 Petersen[4] demonstrated reflecting light beams by small cantilevers driven by electrostatic force. The cantilevers were 100 μm in length, 25 μm in width and 0.5 μm in thickness. They could be deflected by electrostatic force (20 V across 12 μm) over angles of 50°, at high frequencies ($\geq$ 40kHz). Operation cycles of over 10^{10} times were achieved without breakage or any noticeable fatigue. A 16-element light modulator array composed of the cantilevers was used in a small display system. A line of He-Ne laser light was reflected by cantilevers, apertured, focused and scanned by a galvanometer. Though the concept was well advanced, the device was not commercialized because the technology was not matured at that time. Assembly and alignment of the laser, slits, cantilevers, lenses and galvanometers was necessary.

Thanks to technological development, we can imagine the advanced system depicted in Fig. 8.1. Semiconductor lasers, lenses and movable mirrors are all integrated in the system, which has following advantages: (1) miniaturization of the total system, (2) high performance based on parallel processing, and (3) elimination of alignment.

One may notice that the above advantages are in good coherence with key features of MEMS explained in the previous chapter. An array of movable micromirrors on top of a video randum access memory (RAM) has been

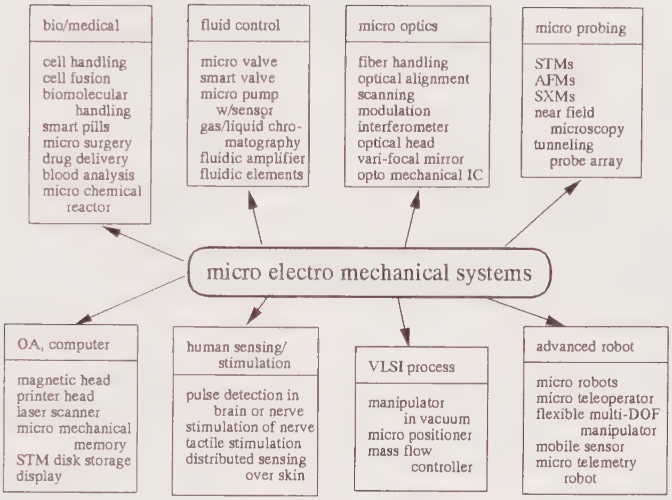

bio/medical
cell handling
cell fusion
biomolecular handling
smart pills
micro surgery
drug delivery
blood analysis
micro chemical reactor
fluid control
micro valve
smart valve
micro pump w/sensor
gas/liquid chromatography
fluidic amplifier
fluidic elements
micro optics
fiber handling
optical alignment
scanning
modulation
interferometer
optical head
vari-focal mirror
opto mechanical IC
micro probing
STMs
AFMs
SXMs
near field microscopy
tunneling probe array
micro electro mechanical systems
OA, computer
magnetic head
printer head
laser scanner
micro mechanical memory
STM disk storage
display
human sensing/ stimulation
pulse detection in brain or nerve
stimulation of nerve
tactile stimulation
distributed sensing over skin
VLSI process
manipulator in vacuum
micro positioner
mass flow controller
advanced robot
micro robots
micro teleoperator
flexible multi-DOF manipulator
mobile sensor
micro telemetry robot

Figure 8.1: Concept of integrated optical MEMS.

fabricated for a projection display. Millions of mirrors are integrated with control electronics [complementary metal-oxide semiconductor (CMOS) processed static RAM (SRAM)]. All the mirrors can be controlled within 100/μs. This system represents the emergence of MEMS-based optical apparatus. Intensive research is under way to realize integrated optical MEMS.

All the mechanical work can be completed within a microsystem for optical applications. Noncontact interfacing between the microsystem and the external world is possible. In addition, most of the optical components, such as lasers, photodetectors, mirrors and optical waveguides are fabricated by semiconductor processes which are compatible with micromachining processes in many cases. Therefore, MEMS have a great potential in optical applications.

Recent developments of optical MEMS are categorized into three groups:

1. light beam control: MEMS devices to control optical signals by mechanical movement;

2. optical sensing: microsensors to detect the change of physical quantities through optical signal (intensity, phase, polarization, etc.); and

3. microstructures actuated by optical energy.

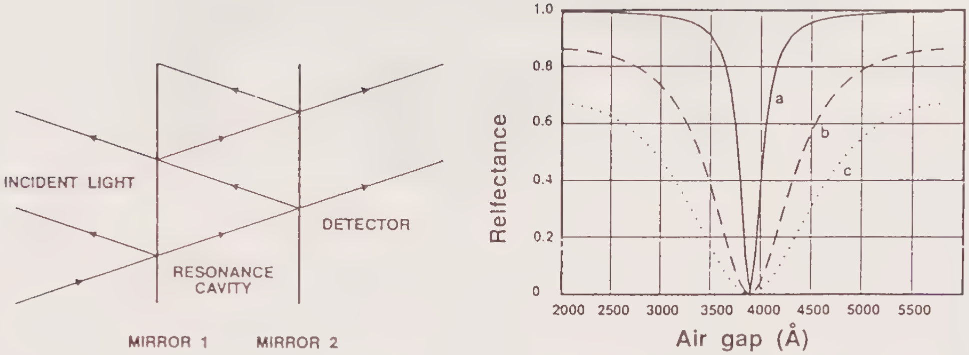
The applications and the examples of MEMS devices in each category are summarized in Table 8.1. Because of the wide range of applications, this chapter mainly focuses on the applications marked with an asterisk in the Table.

Table 8.1: MEMS for optical applications.

Applications	References
(1) MEMS to control optical signals	
Wavelength Filters*	[6] [7] [8] [112] [132] [133]
Wavelength-Tunable Lasers*	[9] [10] [11] [12] [13] [18] [19] [22] [23] [24]
Wavelength-Tunable Photo diodes*	[25] [26]
Modulators*	[4] [6] [27] [28] [29] [31] [32] [33] [34] [35] [36] [37] [38] [134] [135]
Displays*	[5] [39] [87] [88] [89]
Fiber Optic Switches*	[40] [41] [43] [44] [45] [46] [47] [51] [52] [53]
Waveguide Switches*	[52] [54] [55] [56] [57] [58] [59] [60] [61] [62] [64] [65] [67] [68] [69]
Scanners*	[70] [71] [72] [73] [74] [75] [76] [77] [78] [79] [80] [136]
Choppers and Shutters*	[81] [82] [83] [84] [86]
Optical Alignment and Packaging†	[41] [50] [75] [79] [103] [107] [108] [109] [111] [113] [114] [115] [116] [131] [137] [138] [139] [140] [141] [142] [143] [144] [145] [146] [147] [148] [149] [168]
Spectrometers*	[100]
Microlenses*	[108] [110] [118] [119] [120] [121] [122] [123] [124] [125] [126] [127] [128]
Dynamic Focusing Mirrors*	[129] [130]
LIGA for optical elements*	[100] [101] [102] [103]
(2) MEMS for optical sensing	
Near-Field Optical Microscopes*	[90] [91]
Optical Encoders and Displacement Sensors*	[92] [93] [94] [95] [96] [158]
(Optical) Storages*	[97] [98] [99]
Sensors on Waveguides*	[95] [96] [155] [156] [157] [158] [159] [160] [161] [162] [163] [164] [165]
Fiber Optic Sensors	[150] [151] [152] [153] [154]
(3) Optically energized microstructures	
Laser-Manipulation	[166] [167]
Laser-Assisted Deposition, Patterning, and Etching	[169] [170] [171] [172] [173]
Stereo Lithography	[174]
Photostrictive Actuators	[175] [176]
Laser-Excited Oscillators	[18] [177]

*: Application described in this chapter.

†: Some of the devices are referred to in this chapter.



(a) (b)

Figure 8.2: (a) Calculation model of the Fabry-Perot interferometer and (b) reflectance as a function of air gap. Mirror reflectance was set to a:0.9, b:0.5, c:0.3. © (Elsevier Science)

8.2 Recent examples of optical MEMS

8.2.1 Applications of MEMS Fabry-Perot interferometers

A Fabry-Perot interferometer is a very simple interferometer composed of two parallel half mirrors. The principle is similar to an antireflection coating or an oil film on water from which we see rainbow colors. By adjusting the distance between mirrors, it is possible to control the reflection/transmission ratio of the Fabry-Perot interferometer for a given wavelength. Micromachining can be easily used to fabricate such controllable interferometers.

The basic operation of the Fabry-Perot interferometer is shown in Fig. 8.2(a) [6]. The two mirrors with low optical absorption are set in parallel and separated by an air cavity. Using the multiple reflection of the incident light to the left-hand side mirror, one can obtain the total reflectance, R, as a function of the mirror reflectance, r, and the wavelength, λ, by

$$R = \left\{ 1 + \frac{(1-r)^2}{4r \sin^2 (2\pi n d/\lambda)} \right\}^{-1}, \tag{8.1}$$

where d is the gap between the two mirrors. (The transmission ratio T can be obtained simply by $T = 1 - R$.) Figure 8.2(b) shows the dependence of R on the gap d at various values of r ($\lambda = 780$ nm).

Jerman et al.[8] developed an adjustable miniature Fabry-Perot interferometer for multifrequency optical communication. A movable mirror was

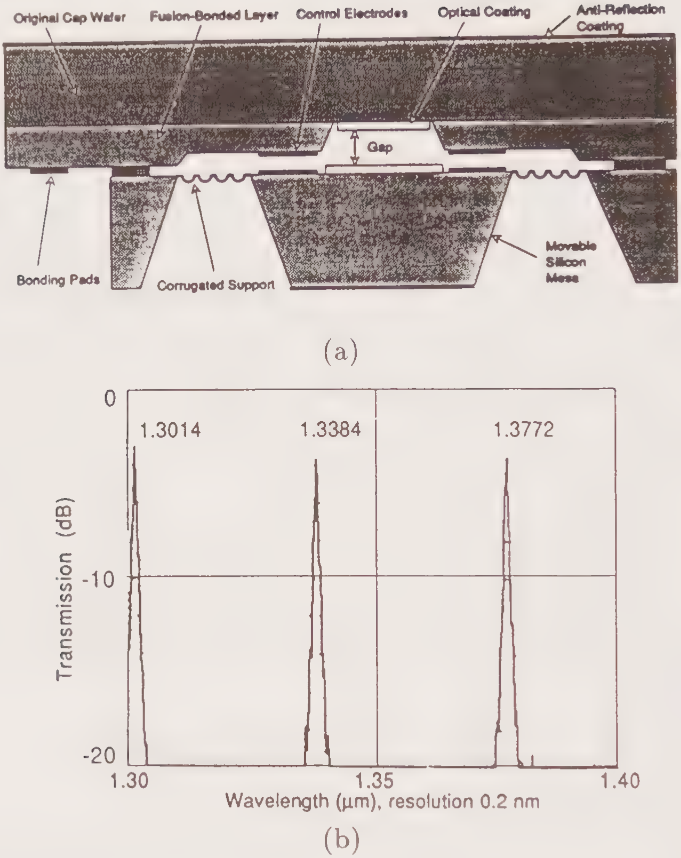

Original Cap Wafer
Fusion-Bonded Layer
Control Electrodes
Optical Coating
Anti-Reflection Coating
Gap
Bonding Pads
Corrugated Support
Movable Silicon Mesa
(a)
0
-10
-20
Transmission (dB)
1.3014
1.3384
1.3772
1.30
1.35
1.40
Wavelength (µm), resolution 0.2 nm
(b)

Figure 8.3: (a) Silicon micromachined Fabry-Perot interferometer and (b) the transmission spectrum. © (1990 IEEE)

fabricated on a silicon substrate and opposed to another half mirror on a glass substrate [Fig. 8.3(a)]. The movable mirror is suspended by a corrugated membrane from the frame. The mirror is thicker than the membrane to keep flatness during actuation. There are four driving electrodes and four sensing electrodes on the glass substrate. These electrodes are used to change the distance between two mirrors and also keep them exactly parallel; the angle is less than 4×10^{-4} degrees. The transmission characteristics for different wavelengths were measured. Many peaks with regular separation were measured as schematically shown in Fig. 8.3(b). One of them exists at 1.34 nm with full width at half maximum (FWHM) of 0.9 nm. The loss at the peak is 30%. The separation between two peaks is 39 nm. The peak shifted 19 nm, which was equal to half of the separation, when 70 V was applied to driving electrodes.

The following sections discuss the three major applications of Fabry-

Perot interferometers: wavelength-tunable lasers (external cavity lasers), wavelength-tunable photodiode, and optical modulators.

External cavity lasers

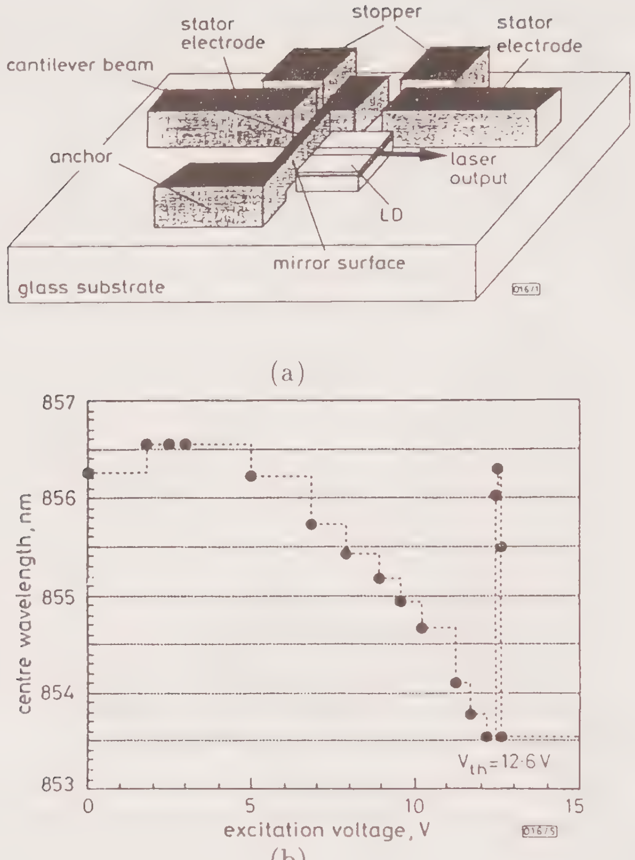
Wavelength-tunable semiconductor lasers have potential applications for wavelength division multiplexed (WDM) optical communications which require wide and continuous tuning range, spectrum narrowing, and the stability of wavelength. However, conventional optoelectronic devices have been limited to a small tuning range of typically only 1 % (corresponding to 10 nm in the near-infrared range), since most of them are based on the modulation of the refractive index through carrier injection or thermal effects. On the other hand, tuning the physical cavity length by micromechanical modulation does not have such limitations. In particular, electrostatic tuning of the cavity resonance with a suspended external mirror enabled tuning ranges of up to 40 nm[9, 10].

There are two approaches in tuning wavelength by using a micromachined external mirror with a semiconductor laser diode: (a) using lateral motion of a mirror with an edge-emitting laser diode and (b) vertical motion with a vertical cavity surface-emitting laser (VCSEL) diode. Both of them employ external mirrors to compose micromechanical Fabry-Perot interferometers controlled by electrostatic force of applied bias voltage.

As an example of tunable edge-emitting laser, Uenishi et al.[11, 12], have reported hybrid tunable lasers on silicon micromachined substrates, shown in Fig. 8.4. They used anisotropic etching of (110) silicon substrates to define a cantilever 1.7 mm long and 8 μm wide. The mirror is the (111) side-wall of the cantilever as it is accurately defined by anisotropic etching and exactly normal to the substrate. The microstructures are transferred onto a glass substrate by using the lost-wafer process to release the structures, followed by the mounting of a laser diode close to the mirror. Uenishi et al. observed the wavelength change from 856 nm to 853 nm with applied voltages of 0 to 12.5 V [Fig. 8.4(b)].

The wavelength tuning was not continuous but stepwise since no antireflection (AR) layer was coated on the device surfaces: without an AR coat, the two resonant cavities (intrinsic cavity of a semiconductor laser and external cavity by mirror) are isolated. In this way the external resonant cavity works as a wavelength filter that picks up one of the resonant modes in the laser. In contrast, by applying an AR coat onto the surfaces, one can see continuous tuning of the wavelength since the outside edge of the laser and the external mirror compose a single resonant cavity. An example will be shown later in this section.

These authors have also proposed nickel electroplating as a fabrication method of external mirrors for edge-emitting laser diodes[13] (Fig. 8.5). The


stator
electrode
stopper
stator
electrode
cantilever beam
anchor
laser
output
LD
mirror surface
glass substrate
D16/1
centre wavelength, nm
857
856
855
854
853
0
5
10
15
excitation voltage, V
V_{th}=12·6 V
D16/5

(a)

(b)

Figure 8.4: (a) Hybrid tunable laser on a (110) silicon wafer and (b) the wavelength tunability. © (IEE)

nickel process is simple compared with the previous one since it uses only resist-molding and electroplating. Since they employed a bidirectional comb drive mechanism, the cavity length could be controlled in both shortening and extending directions. The tunable range of wavelength was thus expanded up to 20 nm through an applied voltage of 20 V.

External mirrors for edge-emitting lasers are made not only of silicon but also of compound materials such as gallium arsenide (GaAs). GaAs is an interesting material for optomechanical devices since it has direct band gap transition, high electron mobility, piezoelectric properties, photoelasticity, and piezoresistive effect[14, 15]. Furthermore, GaAs is considered to be a good material for optomechanical sensors in a high-temperature environment due to its large band gap. A GaAs-based micromachining tech-

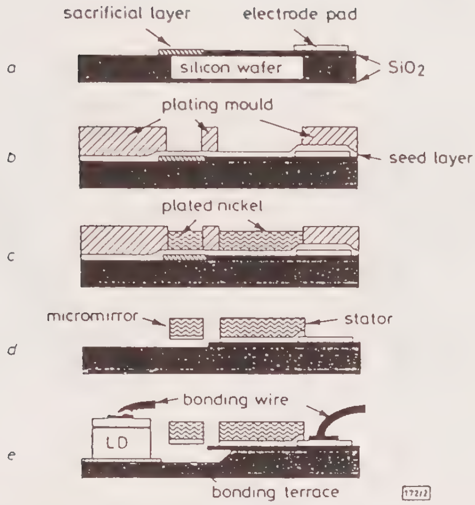
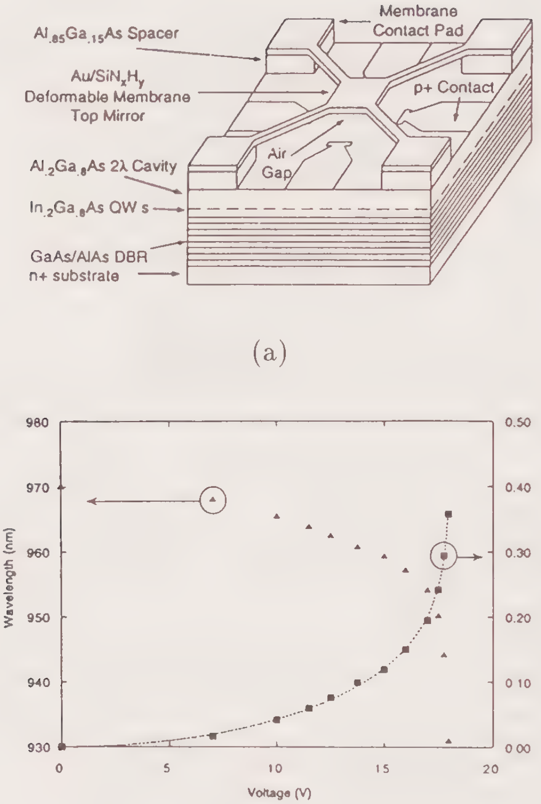


Figure 8.5: External cavity laser with an electroplated nickel microactuator. © (IEE)

nique has been reported by Hjort et al., who employed wet etching to fabricate cantilevers[16], and by Zhang et al., who use a dry process to make microstages[17].

Besides edge-emitting lasers, vertical cavity surface-emitting lasers combined with asymmetric Fabry-Perot modulators (AFPMs) have been attracting attention due to their applicability to high-density optical communication, parallel optical processing, and chip-to-chip optical interconnections[20, 21]. Since they are two-dimensional arrays of integrated lasers, they have good compatibility with the planar IC processing technologies as well as with surface micromachining processes. In both devices, the operating wavelength at the resonant cavity is determined by the layer thickness, which is precisely controllable by epitaxial-layer growth techniques such as molecular beam epitaxy (MBE).

Larson and Harris[10] and Larson et al.[22] developed Fabry-Perot microinterferometers by combining a GaAs-AlAs vertical cavity laser with a suspended movable membrane as shown in Fig. 8.6(a). The structure consists of a bottom mirror, semiconductor cavity, air gap and a deformable-membrane top mirror. The bottom mirror is a 22.5-period GaAs/AlAs distributed Bragg reflector (DBR) with a center wavelength of near $\lambda_0 =$ 975 nm. The GaAs cavity contains three 5-nm InGaAs and 5-nm GaAs quantum wells. The membrane is a three-layer structure made of a 120-nm-thick Au reflector/electrode on top of 380-nm-thick silicon nitride and a GaAs of a quarter wave layer. The air gap thickness is modulated around the initial value of $3\lambda_0/4$ (at zero bias) as a result of electrostatic attraction force from applied voltage. The nitride worked as a phase shift layer, and thus a continuous tuning response was obtained as shown in Fig. 8.6(b). The device exhibited a 40-nm tuning range around the 950-nm center wavelength


Membrane
Contact Pad
Al.85Ga.15As Spacer
Au/SiNxHy
Deformable Membrane
Top Mirror
p+ Contact
Air
Gap
Al.2Ga.8As 2λ Cavity
In.2Ga.8As QW s
GaAs/AlAs DBR
n+ substrate
Wavelength (nm)
980
970
960
950
940
930
Membrane Displacement (μm)
0.50
0.40
0.30
0.20
0.10
0.00
0
5
10
15
20
Voltage (V)

(a)

(b)

Figure 8.6: Vertical coupled-cavity microinterferometer on GaAs and lasing wavelength as a function of tuning voltage. © (1995 IEEE)

for 0 to 18 V applied bias voltage; this corresponded to a vertical membrane motion of 400 nm. The mode suppression ratio was more than 20 dB. The FWHM linewidth was near 1.9 nm, which was not as sharp as that of the edge-emitting lasers because of the short cavity length.

External cavity lasers can be used as sensors to detect physical quantities through the motion of the external mirror on which the strain or stress is applied. Uenishi et al.[18] have monolithically integrated a microbeam resonator with two laser diodes (LDs) and a photodiode(PD) by AlGaAs/GaAs micromachining to create a gas sensor as shown in Fig. 8.7. The microbeam sidewall and the detection LD facet compose a short-coupled cavity laser. The intensity modulated light beam from the excitation LD induces thermal stress in the microbeam to excite resonant oscillation, which is detected by

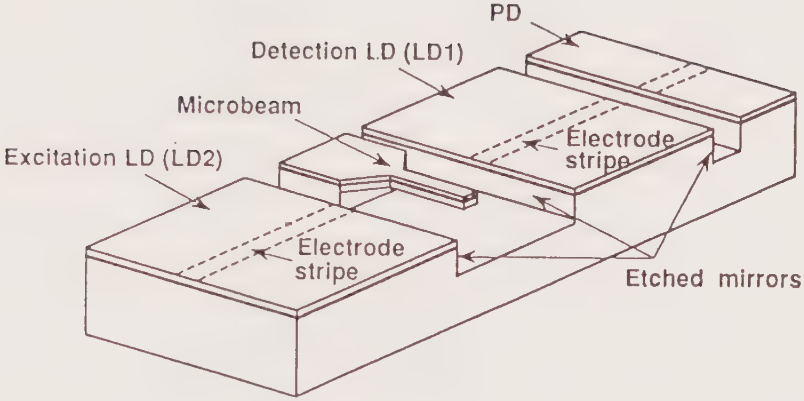


Figure 8.7: Monolithic optical gas sensor in GaAs.

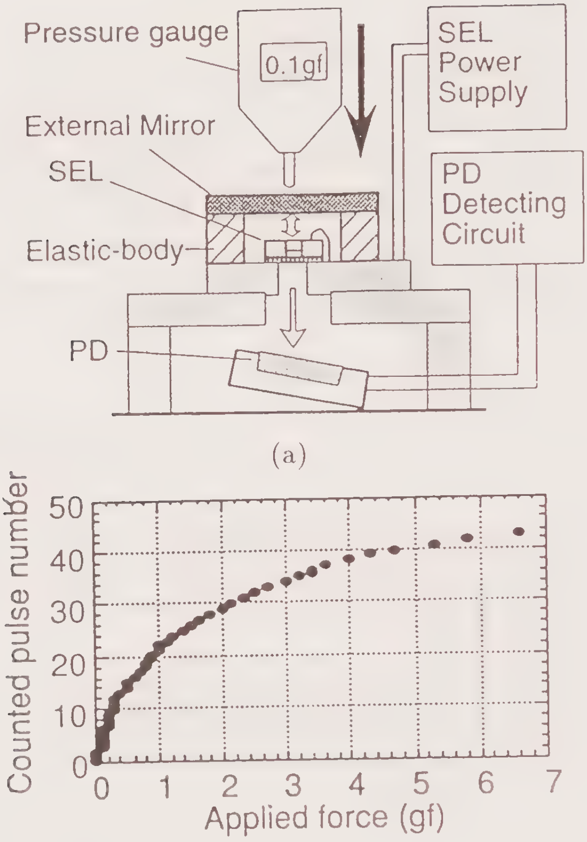
the LD-PD sensor on the other side of the beam. Since the microbeam is very small (110 μm long), the resonant frequency (200 kHz) is high and sensitive enough to detect the mass of molecules (7×10^{-11}g) absorbed on the microbeam.

Yamamoto et al.[23] developed a tactile sensor by using the intensity modulation of the external cavity surface-emitting laser. The device is composed of a laser mounted on a photodiode (silicon), and an external mirror suspended over the laser with a urethane spacer (4.5 mm in diameter and 1 mm in thickness), as shown in Fig. 8.8(a). The external force applied to the mirror changes the cavity length(i.e., the lasing intensity), which is detected with the photodiode. Figure 8.8(b) shows the counted number of the signal peak as a function of the applied force. These authors obtained high sensitivities of 0.01~0.03 gf in a force range less than 0.3 gf.

Ito et al.[24] used a structure similar to the tactile sensor to construct a microdisplacement sensor. In contrast to the previous device, the mirror is separated from the laser by a 0.4-mm clearance, and a piezoelectric bimorph actuator is inserted underneath the photodiode, as shown in Fig. 8.9. The actuator modulates the effective cavity length at a frequency of 1 kHz, so that it is possible to tell the direction of the mirror motion.

Wavelength-tunable photo diodes

Another interesting application of the Fabry-Perot interferometer is wavelength tracking of photodetectors. Wu et al.[25] and Chang-Hasnain et al.[26] have developed the first tunable micromechanical vertical cavity detectors by using a cantilever as a top DBR mirror. They achieved 30 nm continuous tuning for the resonant cavity detector with a 7-V tuning voltage and a 17-dB extinction ratio. Figure 8.10 shows the structure of the tunable detec-


Pressure gauge
0.1gf
SEL Power Supply
External Mirror
SEL
Elastic-body
PD Detecting Circuit
PD
Counted pulse number
0
10
20
30
40
50
0
1
2
3
4
5
6
7
Applied force (gf)

(a)

(b)

Figure 8.8: Optical tactile sensor using external cavity surface-emitting laser diode. © (1995 IEEE)

tor. The top mirror is composed of two sets of DBR reflectors: the movable top reflector is a cantilever of an 18-pair n-doped $Al_{0.6}Ga_{0.4}As$-$Al_{0.1}Ga_{0.9}As$ DBR mirror, and the fixed portion of the top reflector is a 2-pair p^+ doped $Al_{0.6}Ga_{0.4}As$-$Al_{0.1}Ga_{0.9}As$ DBR. The bottom DBR is composed of a 13-pair n-doped AlAs-GaAs mirror grown in an n^+- doped GaAs substrate. By applying a tuning voltage to create electrostatic attraction, the cantilever is deflected toward the substrate to change the resonant Fabry-Perot wavelength.

The center wavelength of Fabry-Perot resonance depends on the voltage difference between the cantilever, V_t, and fixed top reflector, V_f. The detector bias voltage is determined by the voltage difference between V_f and the substrate contact, V_d. The equivalent circuits of the two biasing voltages are

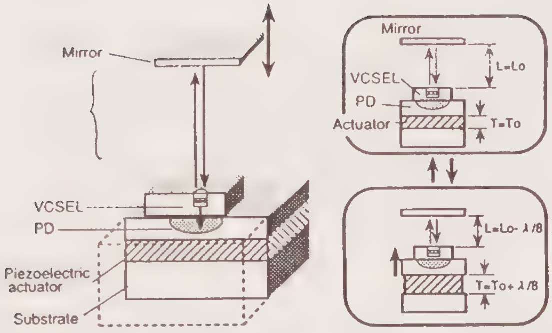

Mirror
VCSEL
PD
Piezoelectric actuator
Substrate
Mirror
L=Lo
VCSEL
PD
Actuator
T=To
L=Lo- λ/8
T=To+ λ/8

Figure 8.9: Structure of microdisplacement sensor based on external cavity laser diode. © (1996 IEEE)

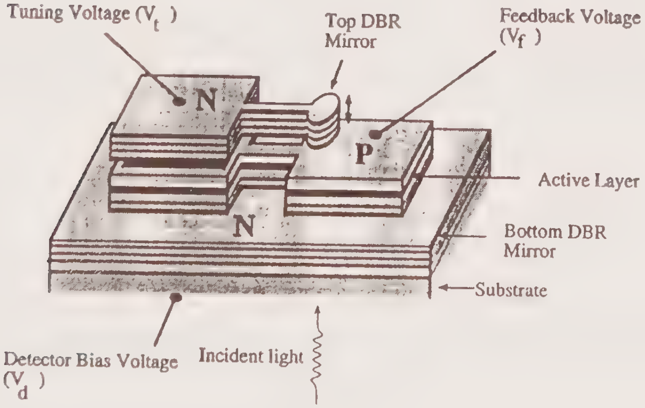

Tuning Voltage (V_t)
Top DBR Mirror
Feedback Voltage (V_f)
N
P
Active Layer
N
Bottom DBR Mirror
Substrate
Detector Bias Voltage (V_d)
Incident light

Figure 8.10: Wavelength-tunable detector with freely suspended cantilever. © (1996 IEEE)

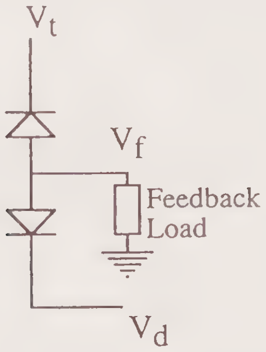


Figure 8.11: Equivalent circuits of the tunable detector and biasing condition. © (1996 IEEE)

shown in Fig. 8.11. Connecting a feedback load to V_f provides the detector with a wavelength tracking mechanism which works in the following way: As the incoming wavelength becomes longer (red shift), the photocurrent from the detector decreases, and the voltage drop across the feedback load is reduced. Consequently, it increases the voltage applied to the cantilever to shift the Fabry-Perot resonance to a longer wavelength, tracking the red shift. On the other hand, when the incoming wavelength becomes shorter (blue shift), the photocurrent increases, and the voltage across the feedback increases to reduce the cantilever voltage, which shifts the Fabry-Perot to a shorter wavelength. Thus the detector could track the wavelength over 25 nm.

Optical modulators

The most practical application of the Fabry-Perot interferometer is an optical modulator used at the subscriber interface of an optical fiber network, as illustrated in Fig. 8.12[27]. In this system the downstream data travel through the fiber splitter, sending the signal to the optical network units (ONU). At the same time, the upstream data from ONUs are sent back to the modulator, which imprints the data on the packet to the central office (fiber-in-the-loop). The advantages of using a micromechanical optical modulator are the maintenance of wavelength coherence of the signal, reliability of the optical device, and temperature insensitivity. Furthermore, it is low cost because no optical transmitters are used in the remote nodes.

Walker et al. [27, 29, 30] and Goossen et al. [28] have developed silicon modulators called mechanical antireflection switches (MARSs), whose schematic is shown in Fig. 8.13[28]. The membrane consists of silicon nitride

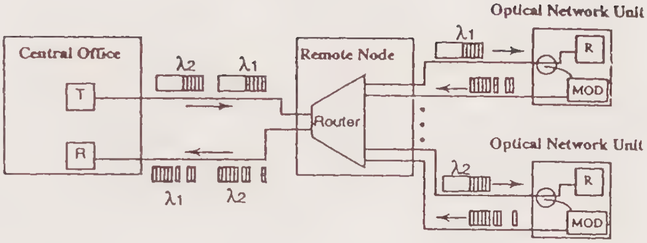


Figure 8.12: Optical fiber network with modulator. Note that no light source is located at the local subscriber. © (1996 IEEE)

(plasma-enhanced chemical vapor deposition) whose thickness and refractive index can be precisely controlled so that it gives an antireflection condition ($\lambda/4$) when the membrane is brought into contact over a wide range from 1.3 to 1.55 μm. The air gap is determined by the silicon oxide thickness m$\lambda/4$. Since the both cavities (membrane and air gap) are coupled, they provide a high reflection at m odd and a high transmission at m even. This device is designed at m $\geq$ 1 to avoid adhesion of the membrane to the substrate.

The temporal response of the device is shown in Fig. 8.14. The rise and fall times under 12 V in atmosphere are about 10 μs (membrane size 100 μm×100 μm). For a smaller size device (20 μm×20 μm), the response decreases to only 250 ns. Contrast ratios of up to 24 dB have been measured. Recently these authors have demonstrated data rates of 3.5 Mbit/s with a 10^{-8} bit error rate in a fiber-to-the-home system.

Various types of silicon micromachined modulators of different structures have been proposed. For example, Anthamatten et al.[31, 32] have developed a modulator with an anisotropically etched hole in a silicon substrate to accommodate an optical fiber by a passive alignment, as shown in Fig. 8.15. Owing to the short light pass, the transmission loss by beam divergence is suppressed. For interchip optical communications, Aratani et al.[6, 7] have proposed a tunable interferometer array which is combined with monolithically integrated photodiodes. Wiszniewski et al.[33] employed SIMOX (separation by implantation of oxygen) SOI (silicon on insulator) wafers to produce micromechanical modulators with a simple fabrication process. The SIMOX technology provides a very thin monocrystalline silicon layer which is suitable for making a diaphragm oscillator because of its low mechanical loss.

As we have seen in these devices, an area larger than a light spot is

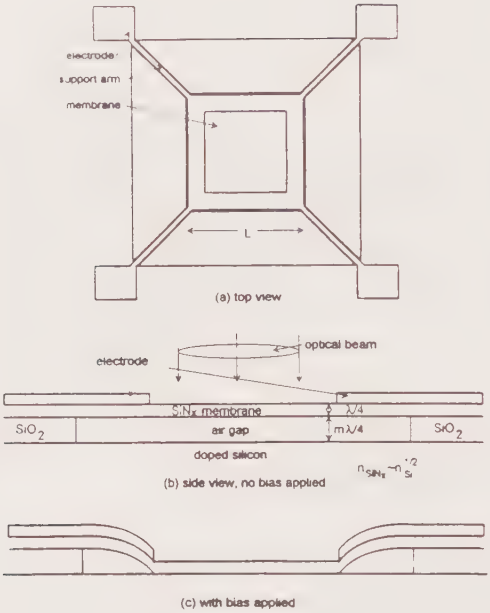

electrode
support arm
membrane
L
(a) top view
optical beam
electrode
SiN$_x$ membrane
$\lambda/4$
SiO$_2$
air gap
$m\lambda/4$
SiO$_2$
doped silicon
$n_{SiN_x} \sim n_{Si}^{1/2}$
(b) side view, no bias applied
(c) with bias applied

Figure 8.13: Mechanical antireflection switch.
© (1994 IEEE)

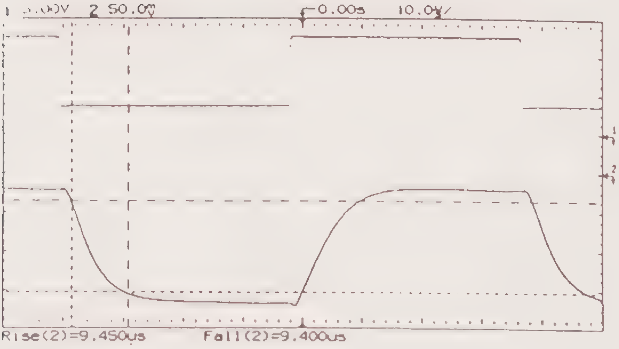


Figure 8.14: Measured response of the MARS device under 12 V in atmosphere. © (1994 IEEE)

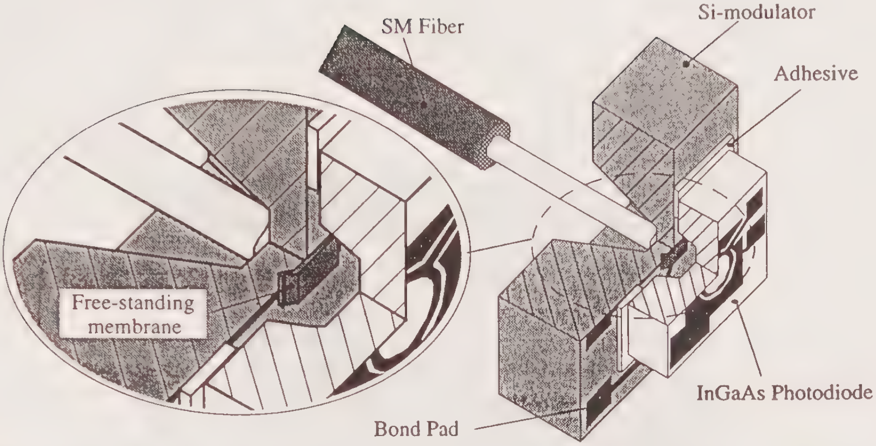
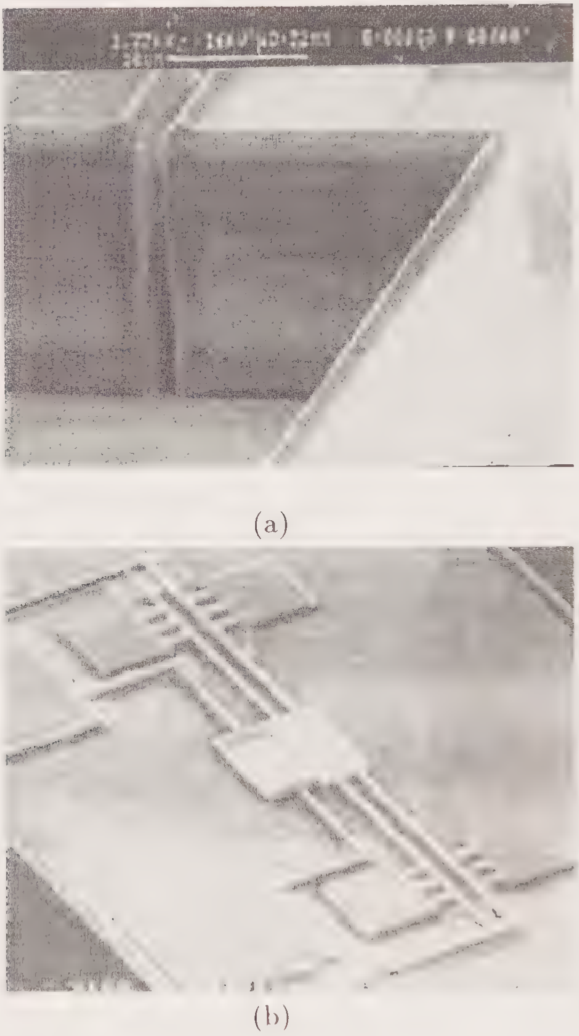

SM Fiber
Si-modulator
Adhesive
Free-standing membrane
InGaAs Photodiode
Bond Pad

Figure 8.15: Silicon micromechanical modulator with an etched hole and an InGaAs photodiode. © (1996 IEEE)

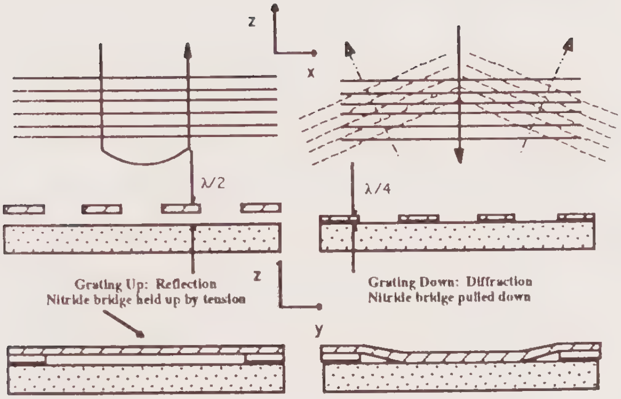
necessary for modulators of high efficiency. Therefore, diaphragms of vertical motion are employed in most of the modulators. However, it is possible to use microstructures with a high wall which are free to move in lateral directions, as an interferometer. MacDonald and Jazairy[34] have developed the SCREAM (single crystal reactive etching and metallization) process to produce an interferometer of high aspect ratio (20 μm high, 2.3 μm gap) and vertical and smooth mirror side walls as shown in Fig. 8.16.



(a)

(b)

Figure 8.16: Fabry-Perot interferometer by SCREAM process.


z
x
λ/2
λ/4
Grating Up: Reflection
Nitride bridge held up by tension
z
y
Grating Down: Diffraction
Nitride bridge pulled down

(a) (b)

Figure 8.17: Operational principle of grating light valve.
© (1994 IEEE)

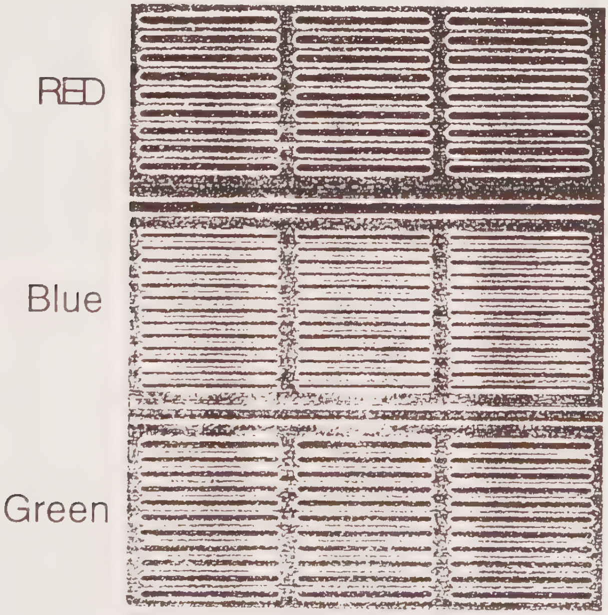

RED
Blue
Green

Figure 8.18: Red, blue and green pixels of the grating light valve.
© (1994 IEEE)

In addition to the Fabry-Perot mechanism, mirror reflection and grating diffraction can be used as an optical modulator. For example, Jaecklin et al.[35] used a 32×32 array of polysilicon torsional mirrors (30 μm×30 μm) as a line-addressable modulator for an application in printing or projecting display devices, and Peter et al.[36] used their torsional mirror as a switching element in a pulsed fiber laser operated at 20 kHz. Solgaard et al.[37, 38] demonstrated a modulator based on an SiN grating, which was suspended over the substrate by $\lambda_0/2$ and actuated vertically through electrostatic attraction. They obtained a bandwidth of 1.8~6.1 MHz and a contrast of 16 dB at switching voltages of 3.2~10 V. Sene et al.[38] used a combination of two layers of polysilicon gratings; the top grating was moved laterally over the other by a thermal actuator for the wavelength tuning.

Fabry-Perot display

Apte et al.[39] utilized an array of Fabry-Perot interferometers as a display. The interferometer is not a simple pair of opposing mirrors but a reflection phase grating consisting of an SiNx microbridge covered with Al. The optical path length between the Al coating and the surface of the substrate is equal to half of the wavelength when no voltage is applied. Therefore, the total path length difference between light reflected from the microbridge and the substrate is equal to one wavelength. These reflections are in the same phase and strengthen each other. The grating works as a flat mirror, as shown in Fig. 8.17(a). When the voltage is applied, the optical path decreases to a quarter of the wavelength. In this case, the total path length difference is equal to half of the wavelength. The reflections are in the reverse phase and the light is diffracted as shown in Fig. 8.17(b). Apte et al. claim that if all the diffracted light in the first order could be collected, an overall efficiency of 84% would be possible. Assuming that a pixel consisting of 10 1-μm pitch gratings is illuminated by white light with 90° incident angle, one may have first-order diffraction beams appearing within 4° around ±14° on the plane perpendicular to both the substrate and the grating. Only the diffracted beam(s) goes to a screen through slit(s). By changing the pitch of the grating, color display can be obtained. Only particular parts of the spectrum, corresponding to red, green, and blue, are allowed to pass through slits. The contrast ratio decreases to 36:1 for red, 59:1 for green and 6:1 for blue. These authors have fabricated pixels of 25 mm by 25 mm consisting of 12 1-μm line pairs. The scanning electron microscope (SEM) picture of the pixel is shown in Fig. 8.18. The operating voltage depends on the design and ranges from 11 to 40 V. The switching speed is 20 ns.

Table 8.2: Comparison of optical switch performances.

Switch	Medium	Insertion Loss	Wavelength Independence	Polarization Independence	Latch	Cost
Fiber-Ribbon Positioner	Optical Fiber	Good	Insensitive	Insensitive	Latchable	High
Micromechanical Switch	Waveguide	Good	Less sensitive	Less sensitive	Latchable	Low
Micromechanical Switch	Optical Fiber	Good	Insensitive	Insensitive	Latchable	Low
Electro-Optical Effect	Waveguide	Acceptable	Sensitive	Sensitive	Battery	High
Thermo-Optical Effect	Waveguide	Good	Less sensitive	Less sensitive	Battery	Acceptable
Laser Diode Gate	Waveguide	Good	Sensitive	Sensitive	No Latch	High
Liquid Crystal	Free Space	Acceptable	Sensitive	Sensitive	Battery	High
Hologram	Free Space	Poor	Sensitive	Sensitive	Battery	High

8.2.2 Micromechanical optical switches

In the development of optical fiber networks, a large quantity of optical switches are necessary for network safety, maintenance, and reconfiguration of the data transmission system. While high-speed optoelectronic devices are widely used in the trunk lines, slow-speed switches would suffice for most of the branch lines because of their low cost. Furthermore, a large switching contrast, small crosstalk, wavelength and polarization insensitivities, and low cost are essential to these subscriber systems. Typical characteristics of various optical switches are compared in Table 8.2. In some applications, switches should be small enough to be mounted on a printed circuit board. To meet these requirements, mechanical regulation of a light beam is suitable as a switching mechanism.

Since the conventional mechanical fiber switches need to be assembled with macroscale actuators or positioners that control the switching state physically, the minimum device size has been limited to a couple of centime-

Table 8.3: Examples of micromechanical optical switches.

Medium	Switching principle	References
Optical fiber	Movable fibers	[41] [44] [45]
	Mirror insertion	[46] [47] [50] [51] [52] [53]
	Interrupter insertion	[40]
Waveguide	Movable waveguide	[58] [59] [60] [61] [62]
	Liquid mirror	[54] [55] [56]
	Mirror insertion	[57]
	Evanescent coupling and spoiling	[52] [64] [69] [67] [68]

ters. Therefore, micromachining is expected to reduce the size, and various kinds of micromechanical optical switches have been proposed.

In the early stage of optical MEMS, Hogari et al. [40] proposed an electrostatically actuated on-off switch which employed a metal cantilever to interrupt a light in a small gap of coupled fibers, and Jebens et al.[41] developed a 2×2 bypass switch using optical fibers actuated by a Joule-heated shape memory alloy.

Several types of micromechanical fiber switches and waveguide switches have been developed. These devices are further categorized into groups based on switching principles, as shown in Table 8.3.

Micromechanical optical fiber switches

One of the most important requirements for optical switches is self retention of the switching state (without flowing power) to ensure security in case of power failure. However, this sometimes makes the switch complicated since it requires a latching mechanism by means of mechanical structures. For example, most of the conventional fiber optic switches[42] consist of a set of optical fibers packed in metal or plastic blocks, as shown in Fig. 8.19, and one of them slides the coupling surfaces to change the switching state.

As a micromechanical version of the fiber optic switch, Nagaoka et al.[43, 44] developed a way to latch fiber switches by a combination of permanent magnets and a coil. Figure 8.20(a) shows a schematic view of the 1×2 fiber optic switch. The fiber core (left-hand side, 125 μm in diameter) has

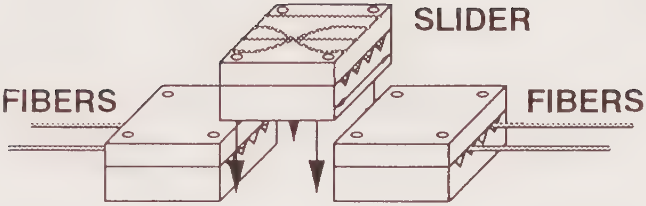
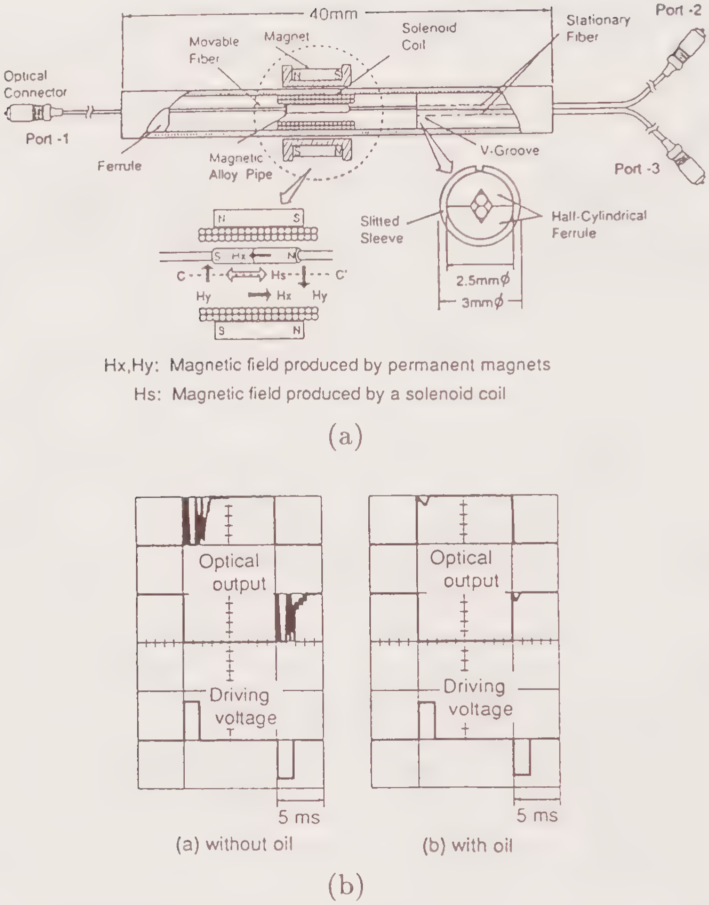


Figure 8.19: Conventional optical fiber switch.

been inserted into a microtube of electroplated FeNi, and is aligned to the output fibers in the ferrule. The solenoid coil located outside of the package attracts the movable fiber back and forth to change the switching state. Every time the coil is activated, the magnetic tube on the core is magnetized and latched by the external permanent magnet. Therefore, no power is necessary to retain the switching state but a short pulse of current (15 mA at 0.6 V) would suffice to toggle the switch. These authors obtained a switching time of 2 ms and crosstalk less than −70 dB. Since the insertion loss was small enough (0.34 dB), these switches can be connected in an N×M matrix configuration. As Fig. 8.20(b) shows, operation in an index-matching oil reduced the chattering caused by the fiber bouncing.

Field et al.[45] developed a 1×2 thermally actuated optical fiber switch, which is shown schematically in Fig. 8.21. The switch is composed of two nickel thermal actuators that move the input optical fiber to align with one of two output fibers. Fibers are precisely held in an anisotropically etched trench. Since the actuating cantilever is composed of two beams of different widths and thus has a difference in resistance, it gives a lateral bending motion through the difference in thermal expansion caused by Joule heating. The actuator can move 250 μm with no load and produce 0.03∼0.5 gf at zero displacement. The input power is 200∼400 mW. Field et al. measured a coupling loss of −5.8 dB while calculation predicts a minimum loss of −0.4 dB. The crosstalk was −66.5 dB. They suggested that latching would be possible by holding down the fibers through electrostatic force.

In these devices the stripped fibers should be long enough to achieve optical coupling with a small loss and with a small physical force of actuation. However, long fibers also determine the minimum size of the device. If a traveling light is redirected by a mirror in a small free space between the


40mm
Movable Fiber
Magnet
Solenoid Coil
Stationary Fiber
Port -2
Optical Connector
Port -1
Ferrule
Magnetic Alloy Pipe
V-Groove
Port -3
Slitted Sleeve
Half-Cylindrical Ferrule
2.5mmϕ
3mmϕ
N
S
Hx
C
Hs
C'
Hy
Hx
Hy
S
N
(a)
Optical output
Driving voltage
5 ms
Optical output
Driving voltage
5 ms
(b)

Hx,Hy: Magnetic field produced by permanent magnets
Hs: Magnetic field produced by a solenoid coil

(a) without oil

(b) with oil

Figure 8.20: (a) 1×2 optical fiber switch by magnetic latching and (b) the switching waveform. © (1991 IEEE)

coupling fibers, the device size can be made much smaller.

Dautartas et al.[46] have developed an optical cross-connector with a silicon-based moving mirror which is inserted into the coupling gap by a magnetic actuator. Figure 8.22 shows the exploded view of the switch package. Four fibers are aligned in a cross on the V-grooves in a silicon substrate. Spherical lenses with an antireflection coat are also aligned with the fibers by the anisotropically etched holes. The mirror is also an anisotropically etched (100) silicon and is coated by sputtered Ti-Pt-Au for a high reflectivity. The mirror is held in place by spring tension and actuated by a wire coil with a ferromagnetic plate. The cross and bar states of the switch are made by inserting and removing the mirror in the coupling gap. The switch loss measured is typically 0.7 dB. The response time is less than 10 ms at a

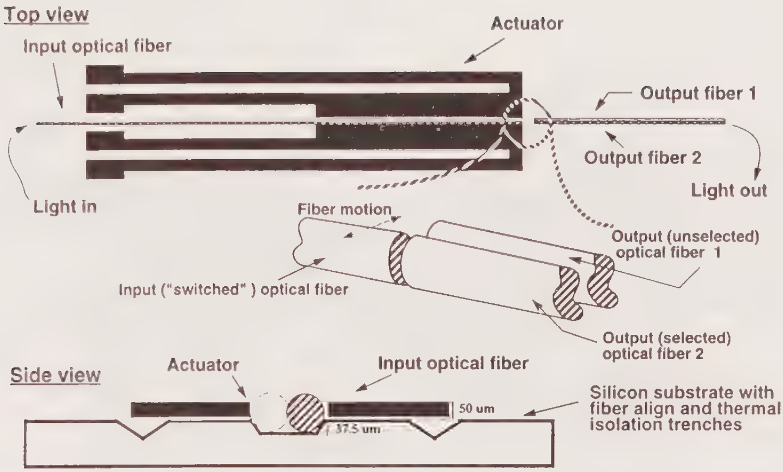

Top view
Input optical fiber
Actuator
Output fiber 1
Output fiber 2
Light out
Light in
Fiber motion
Output (unselected) optical fiber 1
Input ("switched") optical fiber
Output (selected) optical fiber 2
Side view
Actuator
Input optical fiber
50 um
37.5 um
Silicon substrate with fiber align and thermal isolation trenches

Figure 8.21: 1×2 optical fiber switch with thermally actuated nickel beams.
© (Elsevier Science)

driving voltage of 5 V.

For further reduction of the device size, the actuator to insert and remove the mirror can be replaced with a fully integrated MEMS device. Lee et al. [47] developed a cross connector which employed a polysilicon hinged structure[48] as a sliding mirror. Figure 8.23 shows a schematic diagram and an SEM view of the switch. The switch consists of a three-dimensional movable mirror sitting normal to the substrate and four fiber guiding rails. The cross and bar states of the switch are controlled by sliding the mirror into the coupled multimode fibers. Insertion losses of 2.8 dB (cross) and 3.1 dB (bar), and crosstalk of 26.1 dB were obtained at a 1.3-μm wavelength. These authors have recently demonstrated the sliding motion of the mirror by using a scratch drive actuator[49] integrated with the slider[50].

A vertical wall of high aspect ratio created by the LIGA process can serve as a mirror. The LIGA process was developed in Germany in the 1980s and the initials stand for the German words meaning photolithography electroforming and molding. Mohr et al.[51] proposed a 1×2 fiber optic switch with a LIGA mirror as shown in Fig. 8.24. Three multimode fibers are to be accommodated in a "T" orientation, and a mirror in a nonactive state is positioned at the cross-section. Light emitted from the incident fiber is deflected and coupled into the fiber perpendicular to it. When the actuator removes the mirror, the light can pass through the channel and is coupled into the opposing fiber. Since the teeth of the comb-drive actuator have been patterned into a tapered shape, the actuator gives large displacement linearly, up 200 μm at voltages of 70 V.

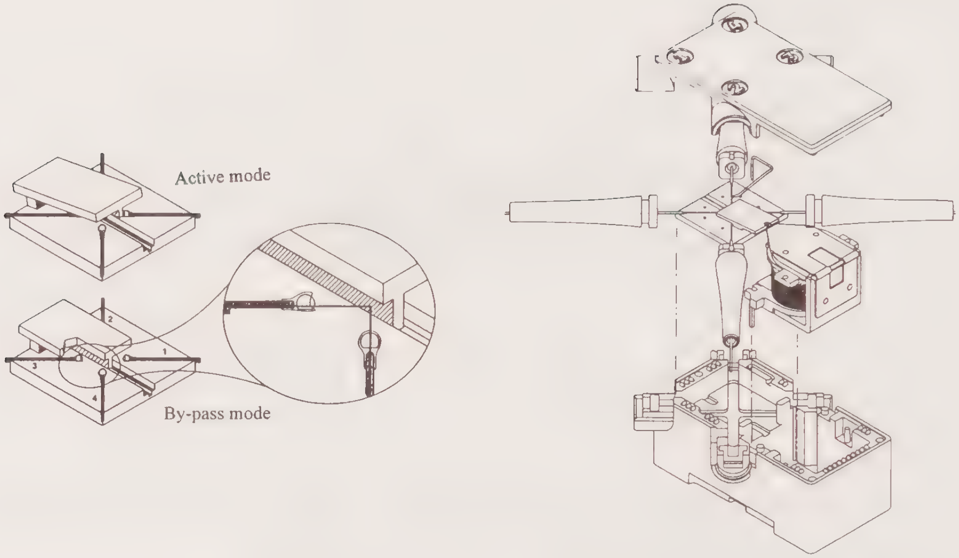
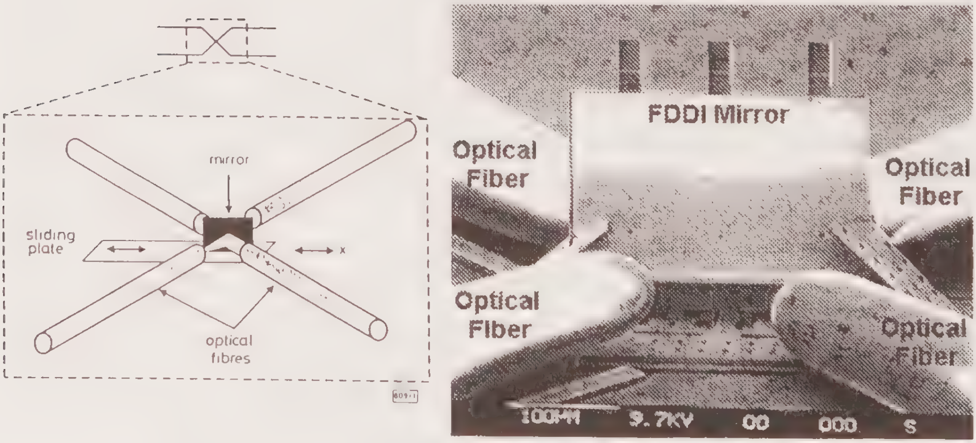

Active mode
By-pass mode
1
2
3
4

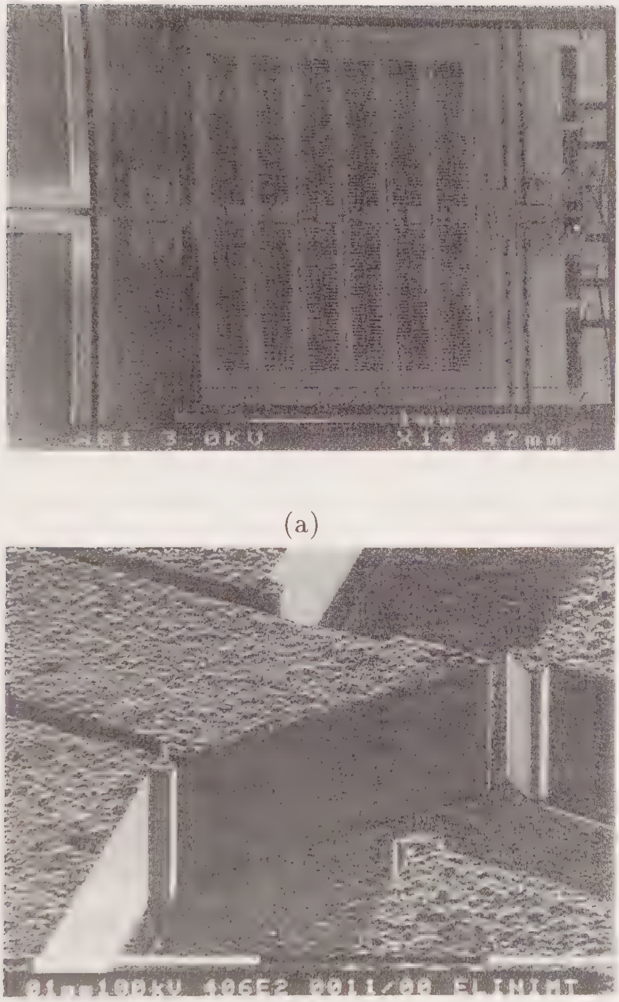
Figure 8.22: Optical cross-connector of a silicon moving mirror inserted by magnetic actuator. © (1992 IEEE)


mirror
sliding plate
x
optical fibres
FDDI Mirror
Optical Fiber
Optical Fiber
Optical Fiber
Optical Fiber
100MM 9.7KV 00 000 S

(a)

(b)

Figure 8.23: Three-dimensional mirror on sliding plate and its application to a surface micromachined free-space fiber optic switch. © (IEE)


1mm
3.0KV X14 47mm
10.0KV

(a)

(b)

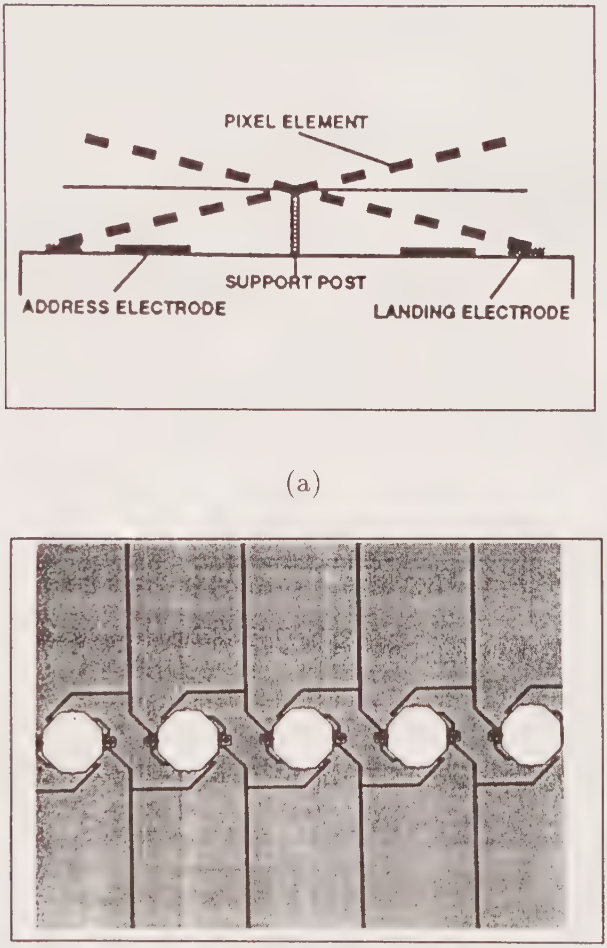
Figure 8.24: SEM view of a fiber optic switch with an electrostatic linear actuator by the LIGA process. © (1993 IEEE)

The advantage of switches based on lateral-motion mirrors is that the coupling gap of the fibers can be set small enough to accommodate the mirrors. Therefore, coupling without lenses is sometimes possible even for the single-mode fibers. However, they need actuators of a large lateral displacement which is comparable to the beam diameter. In contrast, rotating mirrors can swing a light beam with a small angular motion and with a small volume of actuators. However, they need a relatively large coupling distance to shift the beam spot on and off the fiber facet for a sufficient switching contrast.

Boysel et al.[52] have employed their DMD (digital mirror device, see Section 8.2.7) to construct a 4×4 fiber optic switching matrix in a free-space. The switching element consists of an aluminum mirror (50 μm in diameter) supported by two hinges 1 μm wide and 0.1 μm thick with a 5-μm clearance over the substrate. As shown in Fig. 8.25, a voltage of 20 V attracts the mirror by 10° within 10 μs until it comes into contact with a landing electrode. The switching element is constructed by coupling the mirrors with a pair of multimode fibers and lenses. Four of these elements are connected with a star coupler to build a 1×4 switch; and four of these are then assembled into a 4×4 switch, as shown in Fig. 8.26. Boysel et al. observed an insertion loss of −18.7 dB, an extinction ratio of 27.3 dB, and a crosstalk of −28.5 dB.

Toshiyoshi and Fujita[53] adopted micromachined torsion mirrors of large angular motion to regulate light beams in a plane free space. A 2×2 optical switch was developed by integrating electrostatically driven torsion mirrors (Cr-Au evaporated on polysilicon) on a silicon substrate. The dimensions of the mirrors were 300 μm×600 μm and they were supported by thin torsion beams 16 μm wide, 0.4 μm thick and 300 μm long over the through-holes in the substrate. By applying voltage to the mirror and using a counter-electrode substrate, the mirror was attracted inward by 90° to redirect the incident light in a free space. Using collimated-beam optical fibers for optical coupling, these authors constructed a micro free space optical switch on a 3 mm×5 mm chip. The insertion loss was 7.6 dB. Contrast was higher than 60 dB and crosstalk was less than −60 dB. The hysteresis in angle-voltage characteristic of the electrostatic the drive made it possible to hold the mirror with lower voltage than the driving voltage of 100~150 V. No mechanical failure was observed after 40 million cycles at the resonant frequency. A schematic representation of the switch, SEM photographs of the mirror, and switching results are shown in Figs. 8.27 and Fig. 8.28.

Gold is usually deposited on mirror surfaces since it has good reflectivities at infrared and visible wavelengthes. Reflectivities of typical metals used in an IC process are listed in Table 8.4. While chromium is used as a adhesion promoter to gold, aluminum alone can be directly deposited on silicon or silicon oxide. Though silver and copper have good reflectivities, they


PIXEL ELEMENT
SUPPORT POST
ADDRESS ELECTRODE
LANDING ELECTRODE

(a)

(b)

Figure 8.25: Cross-section of a torsion-beam DMD and SEM view.

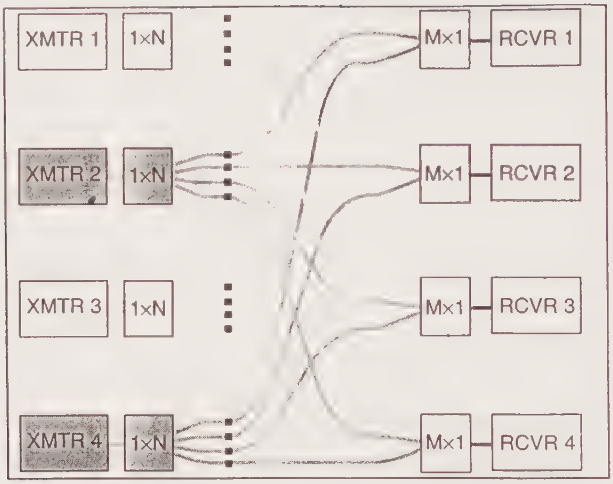


Figure 8.26: Block diagram of 4×4 crossbar switch by DMDs.

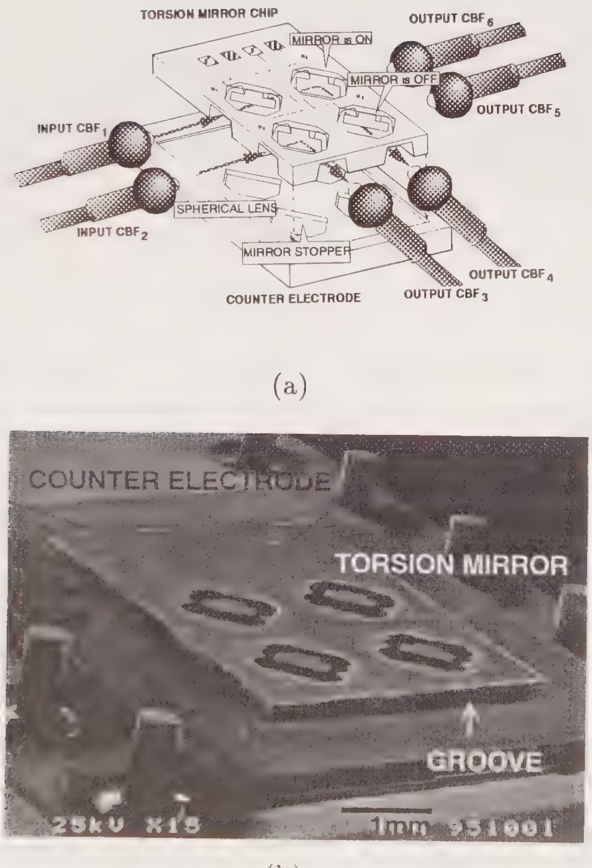
cannot be used as a reflection layer because the surfaces are easily oxidized or sulfurized. It is to be noted that the reflectivity of silicon structures is drastically improved by coating with these metals.

Micromechanical waveguide switches

Switching matrices of large channels of up to 100×100 are usually necessary in an optical main distributing frame, which assigns the optical fibers from customers to the optical subscriber system in a central office. While switches of low duty cycle would suffice, high stability is required for this application. A conventional distributing system consists of thousands of optical fibers whose switching states are controlled by physical coupling and decoupling of fiber connectors.

The function of an optical distributing system can be implemented on a single chip of waveguide switches. However, most solid-state optical switches are not suitable because they need electric power backup to maintain the switching state. Therefore, it is highly desired to keep the switching state physically (self-retaining).

Kanai et al.[54] have proposed a 21×21 switching matrix of self-retaining as a silica waveguide switch. As shown in Fig. 8.29, the device has a vertical groove 15 μm wide, which is formed by anisotropic dry etching at each crossings of the waveguides. Index-matching oil of only 1 nl is filled and removed with a visually controlled nozzle robot to change the switching state: with oil in the groove, the light transmits through the gap and without oil, the light is redirected at the groove surface into the crossing guide. The waveguides are oriented by 60° for the total reflection condition. The inevitable drawback of the switch configuration is the variation in optical


TORSION MIRROR CHIP
OUTPUT CBF 6
MIRROR is ON
MIRROR is OFF
OUTPUT CBF 5
INPUT CBF 1
INPUT CBF 2
SPHERICAL LENS
MIRROR STOPPER
COUNTER ELECTRODE
OUTPUT CBF 4
OUTPUT CBF 3
COUNTER ELECTRODE
TORSION MIRROR
GROOVE
25kV X15
1mm 951001

(a)

(b)

Figure 8.27: Schematic diagram of torsion mirror switching matrix and the SEM view. © (1996 IEEE)

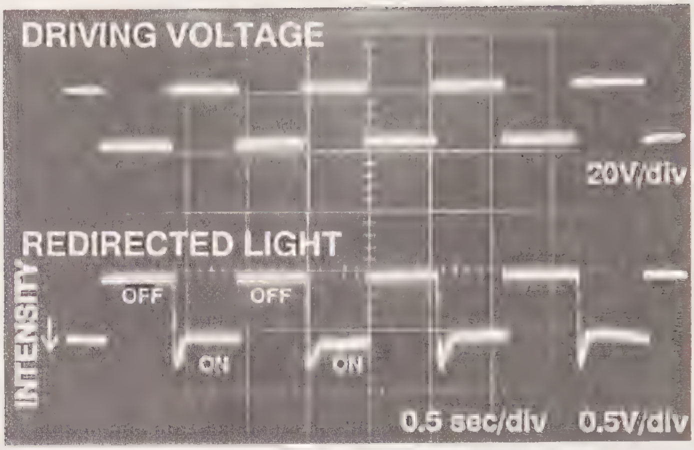


Figure 8.28: Switching curve of a light redirected by a torsion mirror switch.

Table 8.4: Reflectivity of metal.

Metal	Infrared $\lambda = 1.55\ \mu$m ($E = 0.8$ eV)	Visible $\lambda = 650$ nm ($E = 1.9$ eV)
Au	0.993	0.953 (@2.0 eV = 615 nm)
Ag	0.987 (@1.0 eV)	0.944 (@2.0 eV)
Al	0.9778	0.9116
Cu	0.976 (@1.0 eV)	0.943
Ni	0.794	0.659
Pt	0.762	0.678
W	0.710 (@0.82 eV)	0.518
Cr	0.680 (@0.78 eV)	0.644 (@2.0 eV)
single Si	0.306	0.345

Source: CRC Handbook of Chemistry and Physics, (75th ed., David R. Lide, ed.) **12**-117∼140.

$\lambda = \dfrac{1.23 \times 10^3}{E}$ nm.

performance, depending on the route: while there is only one crossing in the shortest route, there are at least 2n − 1 along the most distant route. A typical insertion loss of 7 dB and a crosstalk of 40 dB has been obtained. The switching speed is limited by the positioning time of the robot, $\leq$ 15 s.

Similar types of waveguide switch matrices based on liquid motion have been proposed. Jackel et al.[55] adopted the electrolytic creation and catalytic destruction of bubbles for a bistable switch. A 150-μm-wide slit was sawed into at 45° to the waveguide of silica glass, and a Ti-Pt electrode, which served as a catalyst, was evaporated. The slot was filled with an aqueous solution of acid and covered with a glass to hold the fluid and bubbles. When a pulsed voltage of 20 μs, 7 V was applied between the electrodes, electrolysis occured at the both electrodes and generated oxygen on the cathode and hydrogen on the anode, both of which switched off the transmitted light completely. When the bubbles were merged and mixed well by a large pulse voltage, approximately 1.5 s were needed to reverse the switch through the catalytic reaction. In contrast, with a small pulse voltage, the bubbles were separated and no recombination occured until a second pulse was applied. Thus it worked as a bistable switch. These authors observed a 9~10 dB loss for transmission and 3 dB for reflection of $\lambda = 633$ nm: these losses would be improved by patterning the slots by reactive ion etching.

Sato[56] used a micromirror formed by a mercury droplet immersed in electrolyte. The prototype switch consisted of a resin-molded fiber with a sawed slit 150 μm wide and 100 μm deep. The mercury was driven by electrocapillarity force through the electrically induced gradient in the surface tension of the mercury electrolyte interface, which is schematically shown in Fig. 8.30. He obtained a response time of 20 ms with low driving voltage (0.5~1.0 V) and low power (~25 μW).

Hogari and Matsumoto[57] developed a 2×2 waveguide switch[40] by replacing the optical fibers in their on-off switch[40] with waveguides. A metal cantilever was placed in the waveguide slit and pulled out of it through electrostatic force from an applied voltage of 100 V. Insertion loss of less than 3.1 dB, crosstalk of less than −40 dB, and a response of 2 ms were obtained. By modulating the switching voltage, these workers could suppress the chattering of the switching waveform.

Another approach to waveguide switches is a waveguide structured into a cantilever which is mechanically actuated for optical coupling and decoupling. Bezzaoui and Voges[58] and Hoffmann et al.[59] have patterned waveguides into cantilevers by undercutting the silicon substrate, as shown in Fig. 8.31, and observed a coupling of 3 dB loss for a gap of 2 μm at a wavelength of 633 nm. They used NH_4-H_2O_2-H_2O (AHW) for anisotropic etching of silicon because the standard KOH etching degraded the waveguide layers (PECVD; SiO_2-SiON-SiO_2) due to the nonstoichiometric composition, and because diffused potassium had increased the stress and refractive index of

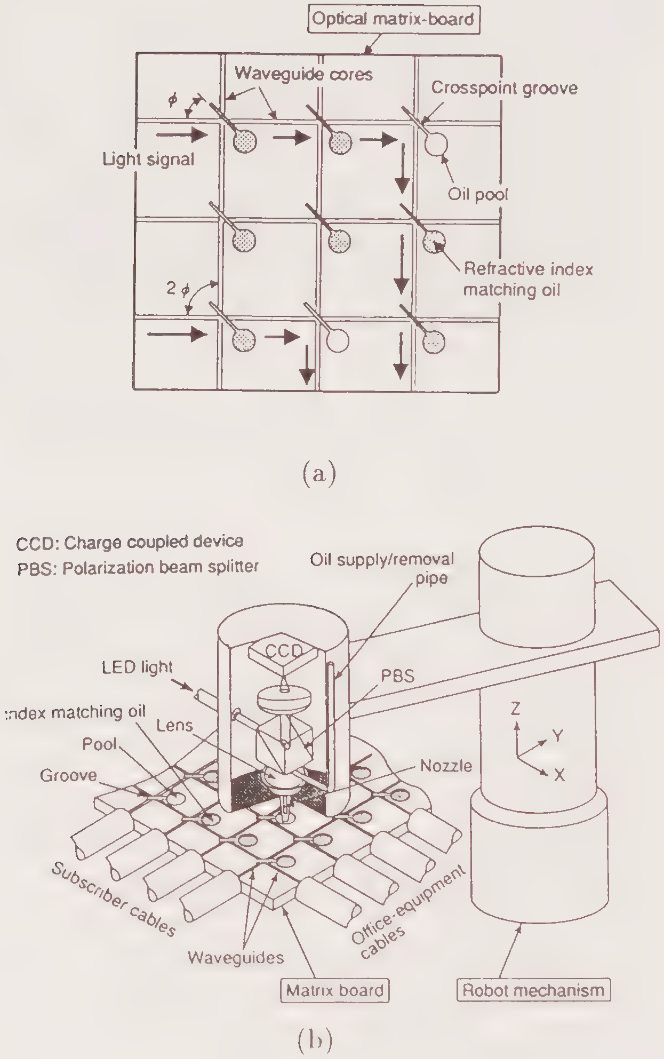

Optical matrix-board
Waveguide cores
Crosspoint groove
ϕ
Light signal
Oil pool
Refractive index matching oil
2ϕ
(a)
CCD: Charge coupled device
PBS: Polarization beam splitter
Oil supply/removal pipe
CCD
LED light
PBS
Index matching oil
Lens
Pool
Groove
Nozzle
Z
Y
X
Subscriber cables
Office-equipment cables
Waveguides
Matrix board
Robot mechanism
(b)

Figure 8.29: Automated optical distributing system based on index- matching oil in a waveguide matrix. © (IEE)

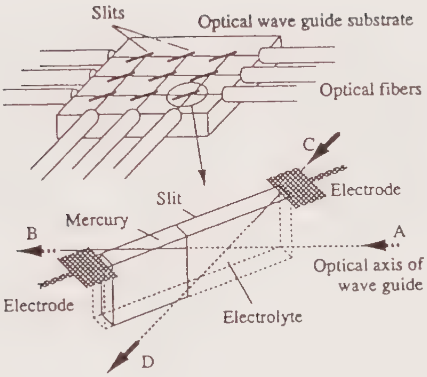

Slits
Optical wave guide substrate
Optical fibers
C
Electrode
Slit
Mercury
B
A
Optical axis of wave guide
Electrode
Electrolyte
D

Figure 8.30: Concept of electrocapillarity optical switch.

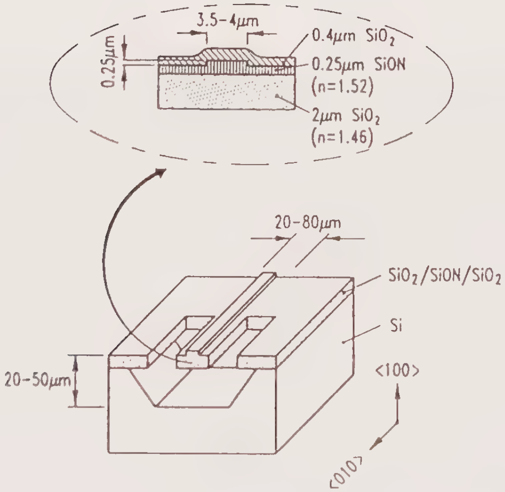

3.5–4μm
0.4μm SiO₂
0.25μm
0.25μm SiON (n=1.52)
2μm SiO₂ (n=1.46)
20–80μm
SiO₂/SiON/SiO₂
Si
<100>
20–50μm
<010>

Figure 8.31: SiO_2-SiON-SiO_2 cantilever with integrated optical strip waveguide. © (Elsevier Science)

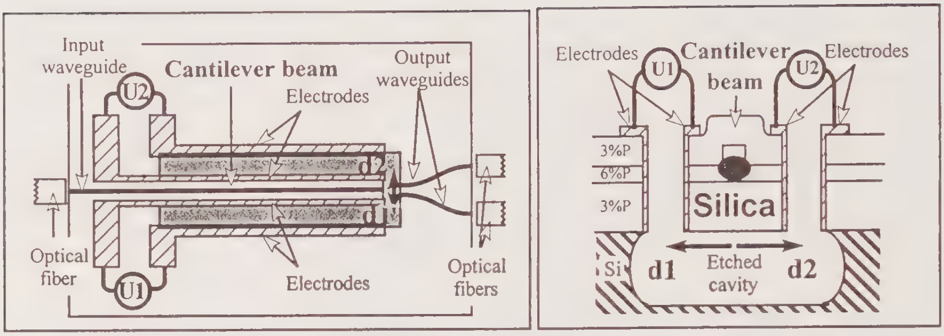


Figure 8.32: Electrostatic 1×2 waveguide cantilever switch of lateral motion. © (1996 IEEE)

the SiON. They put a U-shaped aluminum microheater on the cantilever to induce resonant vibration of 50∼200 kHz, which was detected by the optical coupling.

To simplify the fabrication process of a waveguide cantilever, Eng et al.[60] used an SOI wafer. A cantilever (6 μm wide and a few millimeters long) inserted into two single-mode fibers served as a waveguide for a wavelength of 1.3 μm. With an applied voltage of 10 V, the cantilever deflected to give an on-off switching of 12-dB contrast.

A waveguide cantilever should have a lateral motion in order to construct a switch of two outputs or more. However, a large driving force is necessary for the lateral motion since the width of the waveguide cantilever is usually larger than its height. In a prototype device, Ollier et al. at LETI-CEA[61] used the high wall of the cantilever as an electrode for electrostatic operation, as shown in Fig. 8.32. The cantilever (input waveguide, 25 μm wide, 2 mm long) consisted of three layers of PECVD silica with different phosphorus doping concentrations. When a voltage was applied the cantilever deflected laterally to one of the output waveguides (separated by 20 μm). The total insertion loss of -3.5 dB, crosstalk of -35 dB, and a response time of 0.6 ms were obtained at wavelengths of 1.3 μm and 1.55 μm. Since the total area of the electrodes was limited, the driving voltage was as high as ∼300 V. After these authors had installed a train of electrostatic comb-drive actuators, as shown in Fig. 8.33, the driving voltage was reduced to 28 V[62]. In general, a freely suspended cantilever usually deflects up or down due to the internal stress, leading to an increase in insertion loss. Therefore, Ollier et al. employed a set of compensation narrow beams (5 μm wide and 500 μm long) at the end of the cantilever, as shown in Fig. 8.34, to suppress the offset deflection.

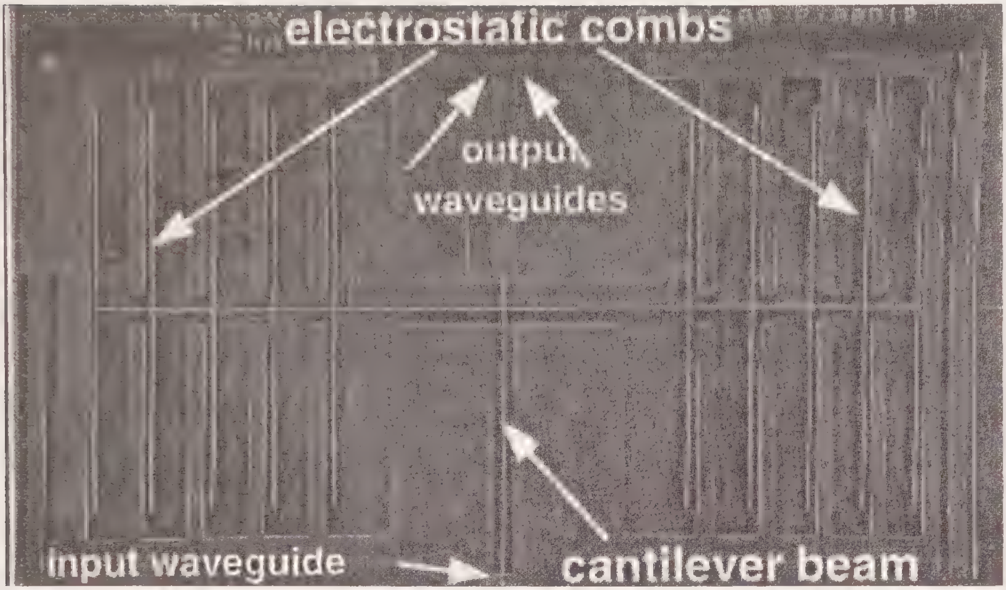

electrostatic combs
output
waveguides
input waveguide
cantilever beam

Figure 8.33: Waveguide switch with comb-drive actuators.
© (1996 IEEE)

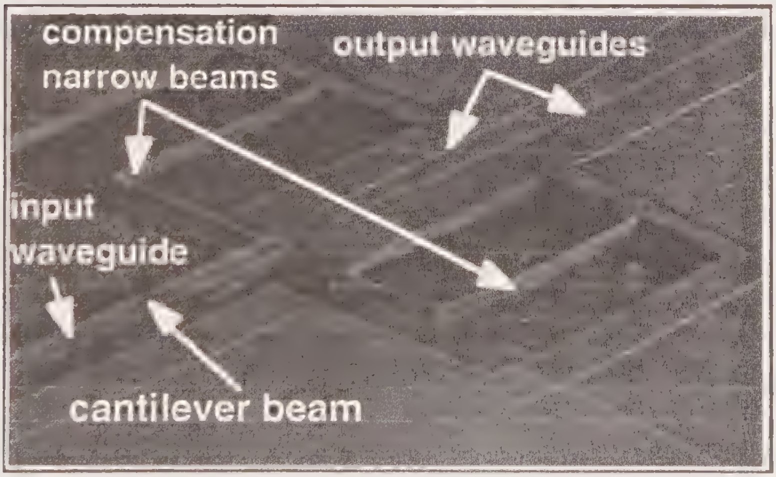
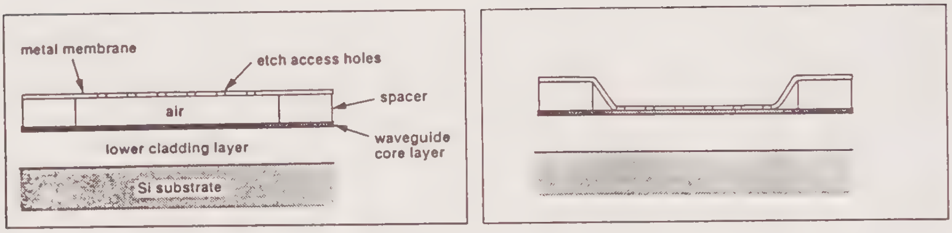

compensation
narrow beams
output waveguides
input
waveguide
cantilever beam

Figure 8.34: Close view of the compensating mechanical beams of the waveguide switch. © (1996 IEEE)



(a) (b)

Figure 8.35: Membrane switch on a PSG waveguide.

Evanescent coupling is an alternative principle for a micromechanical optical switch. When optical fiber cores or waveguides are brought close enough to interact in the evanescent field, where the electromagnetic field of light is leaking out of the medium, the intensity of the light is transferred from one fiber (or waveguide) to the other. Some of the macroscale fiber couplers and switches were based on this effect[63]. Instead of moving waveguides, a membrane or cantilever suspended over the waveguide can be used to change the coupling efficiency[52, 64, 67, 68, 65].

Boysel et al.[52] and Magel[64] used a metal membrane suspended over a phosphosilicate glass (PSG) waveguide as a switching element, as shown in Fig. 8.35. When the membrane is undeflected, the optical mode remains in a guided state. In contrast, when a voltage is applied and the membrane is brought into contact with the guide, the guided light is absorbed in the metal. The calculation of these authors showed that attenuation as high as 30 dB/mm should be possible with an aluminum membrane[52]. By arranging the metal membrane on a directional coupler, a 1×2 switch as shown in Fig. 8.36 is also possible. When no voltage is applied, the membrane is in the noncontact position, and the light is coupled from one waveguide to another (cross state). When voltage is applied and the membrane contacts the output waveguide, the coupling is broken and the switch remains in the barred state. These authors obtained extinction ratios of 6.5 dB in the cross channel and 1 dB in the bar channel. A switching rate of over 1.3 kHz was observed[64].

While a snapping action was used in the earlier devices of Boysel et al. [52] and Magel [64], Pliska and Lukosz [67] and Lukosz [68] controlled the air gap

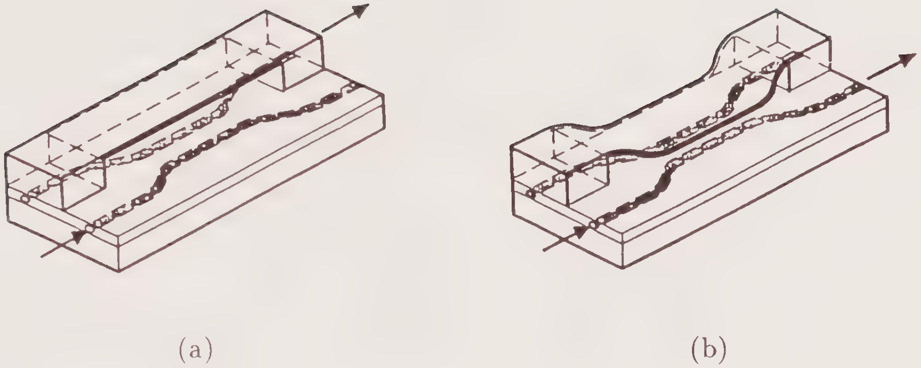


Figure 8.36: Membrane switch with a directional waveguide coupler.

to change the effective index in an analog manner. They used SiO_2-TiO_2 films on Si-SiO_2 substrates as planar waveguides. The modulating elements were bridges or cantilevers of Si-SiO_2 (several mm long and wide) suspended over the waveguide with a small air gap of 300 nm. By squeezing the air gap through electrostatic attraction of the bridge, they observed defocusing, deflection, and focusing of a guided wave, depending on the position incident to the bridge as depicted in Fig. 8.37.

Chollet et al.[69] have proposed the idea of an evanescent wave-coupling switch in which the suspended SiON waveguide is actuated by electrostatic force to change the air gap between the superimposed guides, as shown in Fig. 8.38. When no pressure is applied to the upper guide, the gap is large enough to prevent energy exchange. Thus the light entering the lower guide goes through it. When pressure is applied, the upper waveguide is deflected downward to let the light transfer from the lower guide to the upper guide by the evanescent coupling. Their simulation indicated that the switch showed a through-state with a gap larger than 1 μm and a cross-state with a gap in the range of 0.1 to 0.3 μm for a coupling length of 200 to 400 μm at $\lambda = 633$ nm.

Some of the switching (or modulating) principles, such as a cantilever waveguide and evanescent coupling, can also be used as optical sensors, which will come up in this chapter later. See also Section 8.2.1 for the external cavity laser as a sensing mechanism.

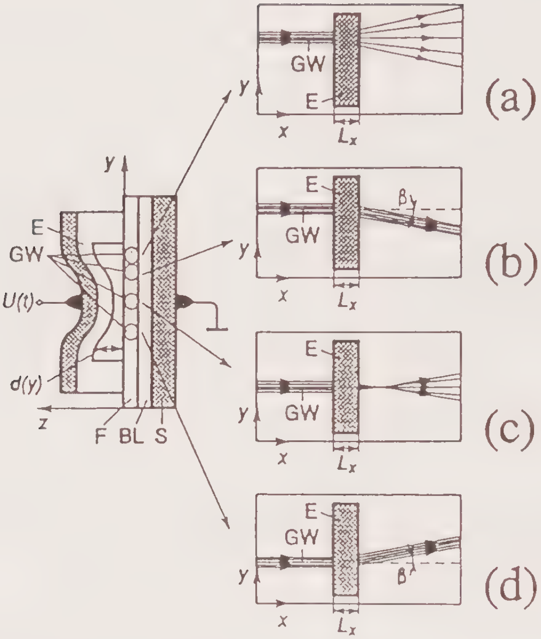

E
GW
U(t)
d(y)
z
y
F BL S
(a)
(b)
(c)
(d)
GW
E
x
y
L_x
β_y
β

Figure 8.37: Electrostatic nanomechanical bridge over planar waveguide. (a) Defocusing, (b) and (d) deflection, and (c) focusing of the guided wave are observed, depending upon the position incident to the bridge.

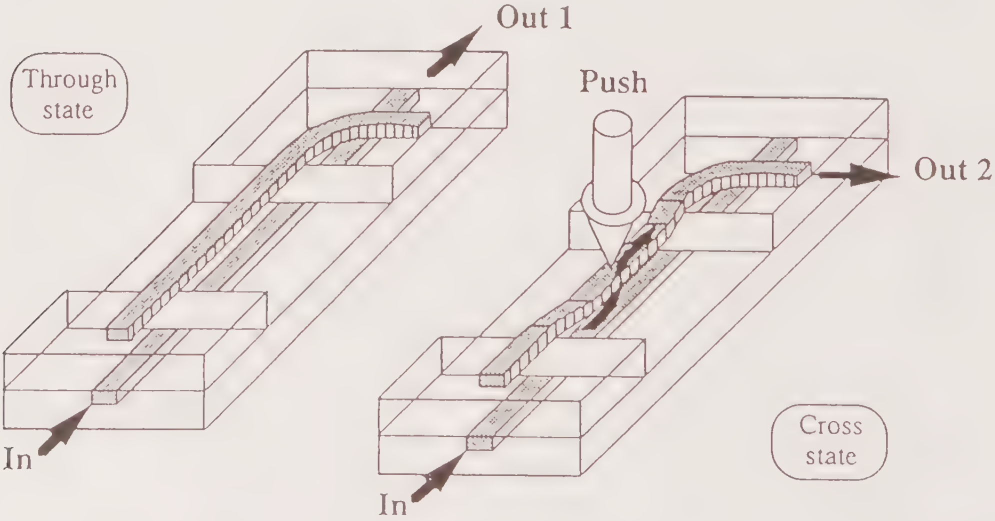

Out 1
Through state
Push
Out 2
In
Cross state
In

Figure 8.38: Optical switch based on evanescent coupling of movable waveguide.

8.2.3 Optical scanners and choppers

Optical scanners are widely used in optical sensors, laser printers, holographic data storage systems and bar-code readers. Conventional scanners are based on polygon mirrors rotated by a motor, galvanometer, acoustic optical modulator, and hologram. Since the scanning angles of polygon mirrors and galvanometers are independent of wavelength, these two devices are mainly used in those applications. To reduce the size and cost of these products, micromachined scanners have recently attracted the attention of industry, and many kinds of prototypes have been developed. Ueda et al. have developed a galvanometer-type optical scanner made by anisotropic etching of quartz and operated with an integrated coil which interacts with the external magnetic field.

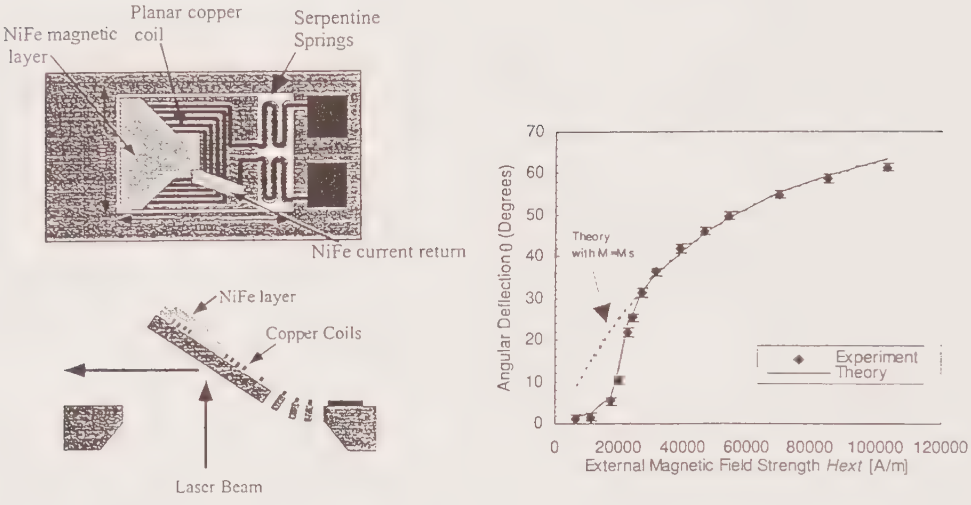
Furthermore, reduction of the size of scanners has opened up a new application such as a microrobot that can investigate the inner wall of tubes by optical or acoustical sensing. For example, Takahata et al.[71] have developed a micromotor fabricated by electrodischarge machining. A micromirror is attached to the shaft of the wobbling rotor to reflect ultrasonic waves used to detect cracks on the wall of a thin tube.

Miller et al.[72] fabricated a millimeter-sized analog scanner, as shown in Fig. 8.39(a), which has scan angle of over 60°[Fig. 8.39(b)]. Deflection was achieved by controlling the external magnetic field and the current through a 30-turn coil, which was integrated on the mirror plate with an 11-μm-thick Permalloy layer. The external magnetic field was used for coarse deflection control (or providing bias deflection) and current in the coil was used to control the fine angle or to cause fast motion. Miller et al. demonstrated the application of this mirror in a holographic data storage system.

Judy and Muller[73] have also developed a magnetically operated scanner, which is a torsional mirror operated through an internal magnetic field. They integrated a plated nickel mirror (450 μm square and 5 μm thick) on polysilicon flexures over a 10-turn coil on a silicon substrate. With a current of 500 mA in the coil, a field of $\sim$5 kA/m is generated to turn the mirror by more than 45°. Electrostatic latching is possible by applying a voltage to the mirror and the substrate.

Kiang et al.[74, 75] used polysilicon hinged microstructure technology to make scanners (Fig. 8.40) driven by electrostatic comb-drive actuators. One of their devices was a 200 μm×250 μm scanning mirror which could scan 12° at the resonant frequency of 3 kHz with a driving voltage of 20 V[75]. The scanner was used with macroscopic optical components, and a bar-code reading operation was demonstrated.

Ikeda et al.[76, 77] fabricated a two-dimensional scanner integrated with photodetectors and piezoresistors. In order to isolate the photodetector from the peripheral integrated circuit on the chip, they suspended the detector



(a) (b)

Figure 8.39: Microscanner controlled by a planar coil in the magnetic field. © (1996 IEEE)

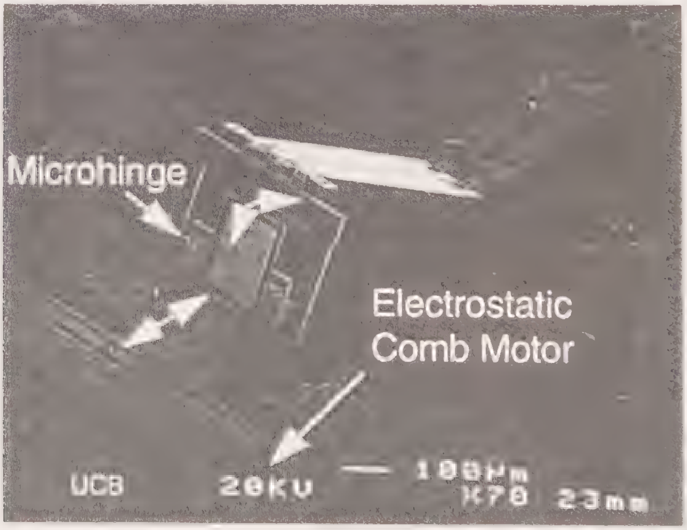
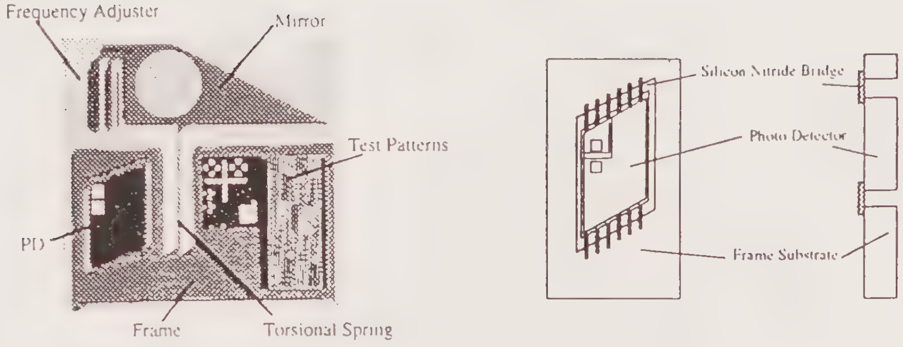

Microhinge
Electrostatic Comb Motor
UCB
20KU
100μm
X70
23mm

Figure 8.40: Electrostatic comb-driven microscanner with a polysilicon hinged structure technology. © (1996 IEEE)



(a) (b)

Figure 8.41: XY scanner by anisotropic etching of silicon with suspended photodetector.

island with silicon nitride bridges as shown in Fig. 8.41. Actuation was done by an attached bulk piezoelectric actuator, which had a displacement expanding mechanism. The scanner had a mirror supported by thin and long beams. The mirror resonated at 577 Hz in bending and 637 Hz in torsion. A laser beam was scanned over 40°×30° and the reflection from the object was detected by photodetectors. The scanning angles were detected by piezoresistors.

Two-dimensional scanning has been also demonstrated by Ohtuka et al.[78] with an orthogonal set of torsional beams supporting a mirror as shown in Fig. 8.42. The frame is made of beryllium copper (9.3 mm×9.3 mm× 0.5 mm) patterned in ferric chloride etchant and piezoelectric bimorph actuators are attached. Two-dimensional scanning of ±30° has been achieved by superimposing driving voltages (40∼100 V) at the resonant frequencies (510 Hz and 916 Hz) of the X- and Y-moments of the mirror.

Mechanically movable gratings can be also used as a scanner. Yasseen et al.[79] have integrated a grating of 1.8 μm pitch on a polysilicon micromotor of 500 μm diameter(Fig. 8.43), and scanned the diffracted light with the electrostatic operation of the motor. The scanning motor was operated at 18 V, and the maximum speed was 1100 rpm. To improve the surface quality, they applied chemical-mechanical polishing to reduce the surface roughness to 20 Å. They demonstrated bar-code scanning without the external lens system. Merat et al.[80] developed a polygon microscanner by electroplating of nickel (20 μm high) on a polysilicon micromotor, as shown in Fig. 8.44.

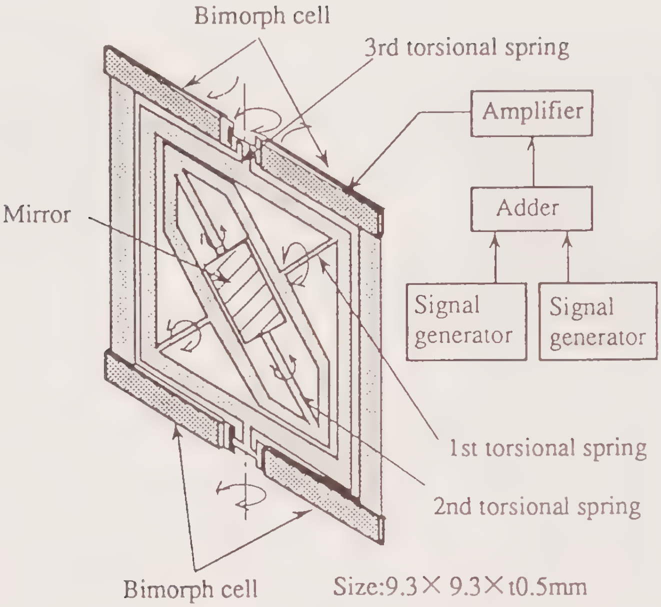

Bimorph cell
3rd torsional spring
Amplifier
Adder
Mirror
Signal generator
Signal generator
1st torsional spring
2nd torsional spring
Bimorph cell
Size:9.3×9.3×t0.5mm

Figure 8.42: Two-dimensional optical scanner with orthogonal torsion beams. © (1995 IEEE)

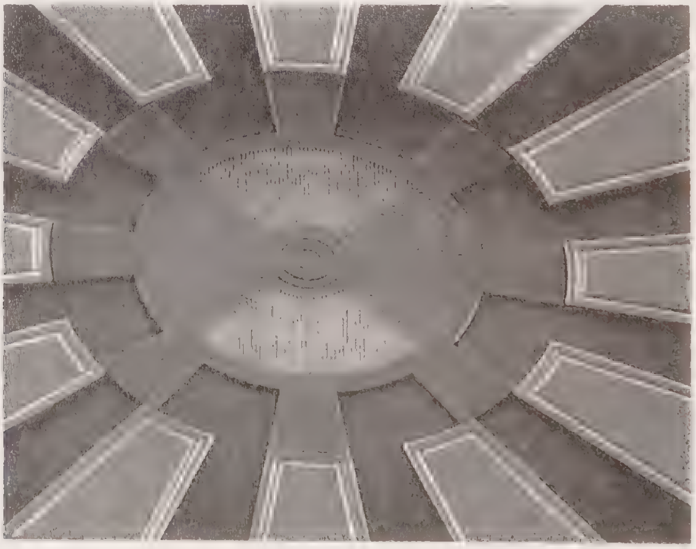
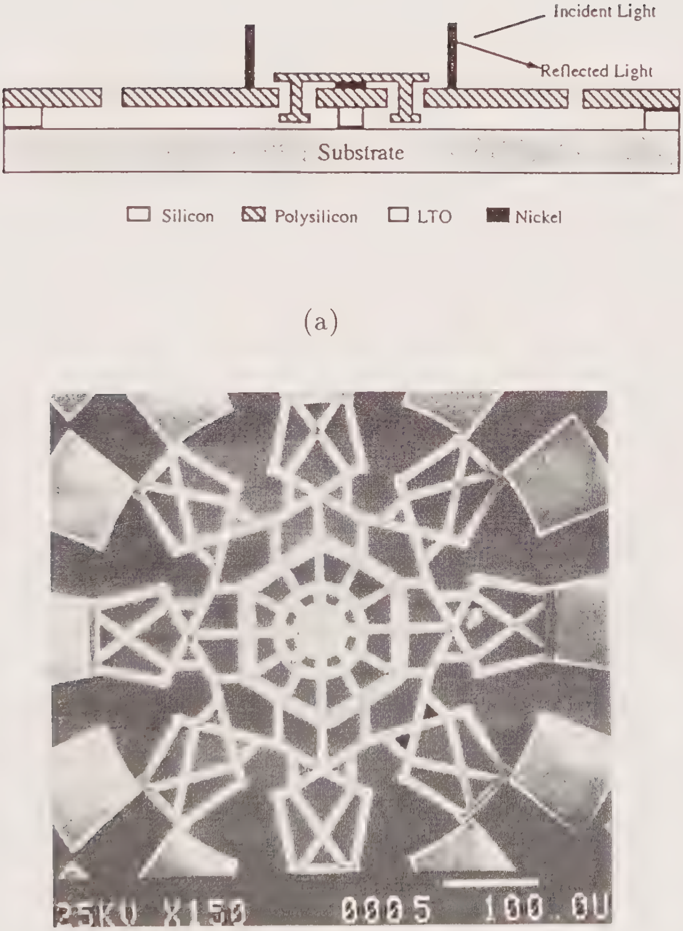


Figure 8.43: Micromotor scanner with polished polysilicon grating of 1.8 μm pitch. © (1995 IEEE)


Incident Light
Reflected Light
Substrate
☐ Silicon
▨ Polysilicon
☐ LTO
■ Nickel
25KV X150 0005 100.0U

(a)

(b)

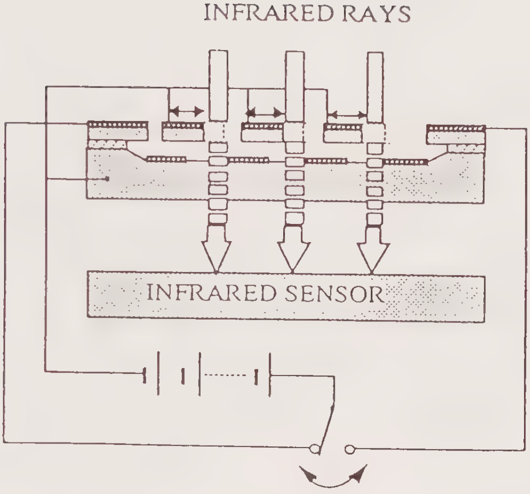
Figure 8.44: Polygon scanner on a micromotor.

Optical choppers (shutters) are usually used in precision infrared (IR) optical sensors of pyroelectric detectors. Since pyroelectric material can detect the rate of temperature change, it needs to be alternately exposed to the light source and a reference body. Some of the macro-scale IR sensors use a motor-driven mirror to switch the light from the source and the reference into the detector intermittently, and the rest of them use motor-driven disks of slits to regulate the light from the source into an ac signal: in the latter case the chopper itself works as a reference body. Compared with these conventional choppers, the micromachined chopper is expected to have advantages from the viewpoints of compact size, low noise, low power consumption, and long lifetime.

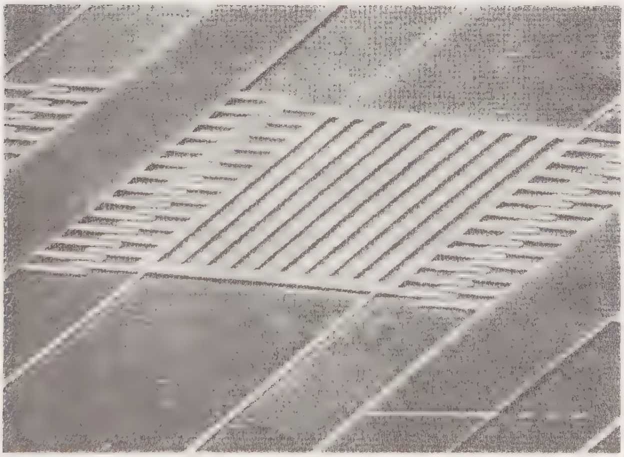
The basic idea of the micromachined optical chopper is depicted in Fig. 8.45(a) [81]. The incident light coming down to the device is regulated into an ac signal by the mechanical motion of the chopper and is detected by the infrared sensor underneath the chopper. Since silicon is transparent in a wide range of infrared wavelengths, the substrate does not necessarily have through-holes. Tabata et al.[81] have developed a chopper by using an SOI wafer, which is shown in Fig. 8.45(b). After the patterning of the suspended chopper (boron-implanted silicon, 1 μm) by reactive ion etching, the sacrificial layer (silicon oxide, 1 μm) is removed. To have a clearance of 10 μm beneath the chopper plate, the substrate is anisotropically etched in tetramethylammonium hydroxide (TMAH). Since the upper silicon is B^{++} doped (4×10^{16} cm^{-2}), it survives in the etching. The barrier metal (Al) to shut out the incident light is formed by vacuum deposition by using the chopper slit as a mask (self-alignment). The chopper plate is driven by the comb-drive actuators on the sides. An optical shutter with a comb-drive actuator has been also proposed by Jaecklin et al.[82]. They formed a through-hole under the shutter by anisotropic etching from the backside of the silicon substrate.

Obermeier et al.[83] have developed a torsional shutter suspended on a through-hole into a silicon substrate, as shown in Fig. 8.46, and improved the fill-factor of the chopper. The shutter (0.8 mm×1 mm in area and 1 μm in thickness) consists of Al, Cr-Au, or doped polysilicon, and is supported by a beam 450 μm long and 10 μm wide. Because of the nonlinearity in the electrostatic attraction force, a voltage of 30 V is required for opening the mirror by $\theta = 30°$, and only 15 V is needed to keep it at $\theta = 90°$(full open).

The standard process of fabricating quartz oscillators can be also used to make a piezoelectrically operated optical chopper. Toshiyoshi et al.[84] developed quartz actuators by anisotropic etching in buffered hydrofluoric acid(HF) and applied them to a resonant-type optical chopper, as shown in Fig. 8.47(a). The through-holes on the chopper plate were used as chopping slits. Figure. 8.47 shows the meandering electrodes (Cr-Au) on the suspension to induce the deflection in a bending mode. The piezodrive suspensions were used both to excite and to detect the oscillation, and the oper-


INFRARED RAYS
INFRARED SENSOR

(a)



(b)

Figure 8.45: Concept of optical micromachined monolithic optical chopper made of SOI wafer. © (1993 IEEE)

ation of stable frequency and amplitude was achieved by a feedback control. A resonant amplitude of 60 μm was obtained under an applied voltage of $\sim$100 V at $\sim$2.7 kHz. These authors also developed a piezoelectric quartz actuator of large stroke by building a suspension folded 20 times to increase displacement[85]. It is noteworthy that the mechanism generated large static displacement under ac voltage operation, through a strong nonlinearity in the oscillation.

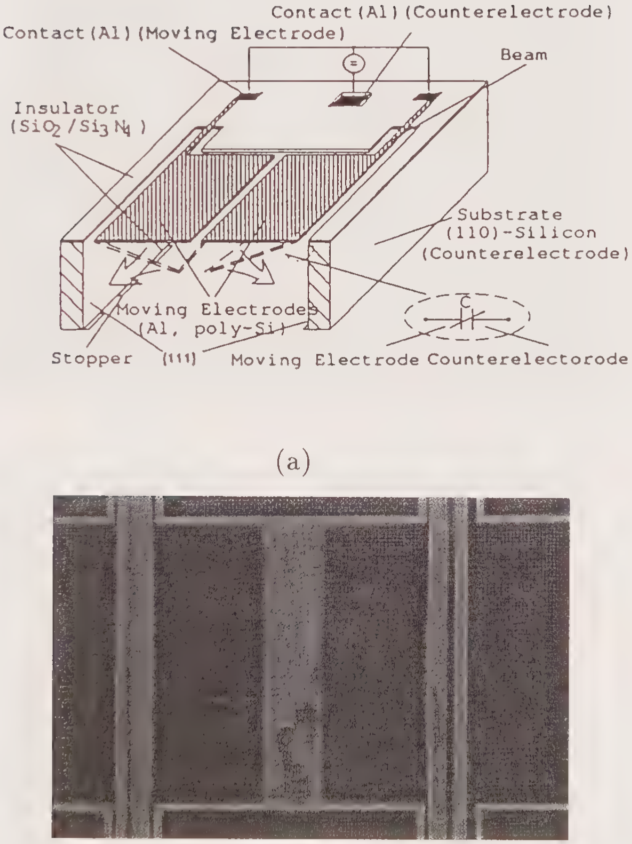
Micromechanical slits for precision optical spectrometers have been developed by Vuilleumier and Kraiczek[86]. A typical spectrometer employs a pair of entrance and exit slits before and after the grating. The slit widths and positions should be adjusted to improve the spectral resolution. These authors developed an adjustable slit by using a monolithic frame made of silicon and combining it with a stepping motor via a gear train, as shown in Fig. 8.48. Several slits of different widths were formed in the silicon plate and placed beneath the aperture plate. They obtained a reproducibility of the slit center position better than 0.01 μm.

8.2.4 Optical elements using three-dimensional polysilicon structures

There are a number of technologies for making three-dimensional structures out of micromachined polysilicon patterns. Pister et al.[48] and Yeh et al.[104] made polysilicon hinges and folded polysilicon microstructures out of the substrate plane. Suzuki et al.[105] used polyimide elastic joints between polysilicon films and also folded them to make an "insect-based" microrobot. Fukuta et al.[106] plastically deformed polysilicon beams into three-dimensional structures by using the Joule heat of electric current.

Recently three-dimensional microstructures have attracted attention because of their potential applicability to free-space micro-optics. The concept of a free-space micro-optical bench (MOB) has been proposed by Wu et al.[107], which is the integration of an optical system such as microlenses, mirrors, gratings, lasers, and detectors onto a single chip. They used polysilicon hinged structures to build optical components such as mirrors and gratings located perpendicular to the wafer surface.

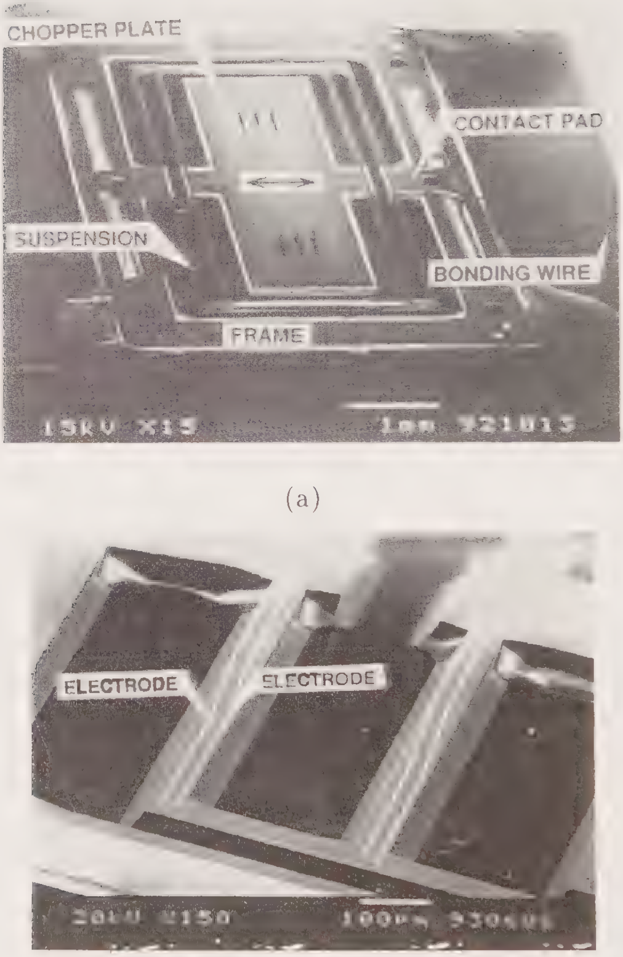
The fundamental hinge fabrication process of two-layer polysilicon with oxide layers is shown in Fig. 8.49[48]. A sacrificial phosphosilicate glass layer is deposited on a silicon substrate, followed by a polysilicon layer by low pressure chemical vapor deposition (LPCVD). After patterning the polysilicon (mirror), another PSG layer is deposited. Contact holes are patterned into the PSG layers and the top polysilicon layer (hinge staple) is then deposited and patterned. The sacrificial oxide layers are removed, and the first polysilicon patterns are free to rotate. Extra beams snapping the structure out of the wafer surface can be also produced in the same manner.


Contact (Al) (Counterelectrode)
Contact (Al) (Moving Electrode)
Beam
Insulator
(SiO2 / Si3 N4)
Substrate
(110)-Silicon
(Counterelectrode)
C
Moving Electrodes
(Al, poly-Si)
Stopper
(111)
Moving Electrode
Counterelectorode

(a)

(b)

Figure 8.46: Electrostatic torsional microshutter on (110) silicon. (a) schematic view and (b) SEM view of the shutter opening 30°. © (1993 IEEE)


CHOPPER PLATE
CONTACT PAD
SUSPENSION
BONDING WIRE
FRAME
15kV X15
1mm 921015
ELECTRODE
ELECTRODE
20kV X150
100µm

(a)

(b)

Figure 8.47: Quartz optical chopper excited through piezoelectricity by using the meandering electrodes on the suspensions.

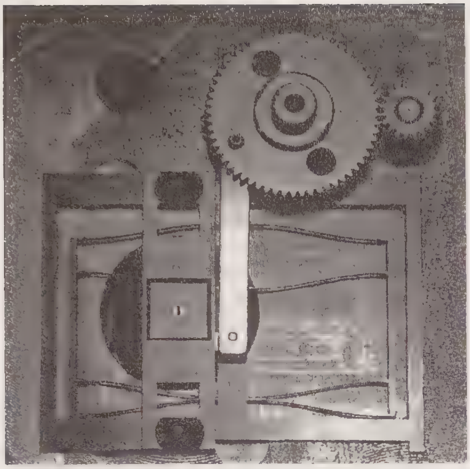


Figure 8.48: Variable slit system composed of silicon plate, driving gears, and connecting bars. © (1995 IEEE)

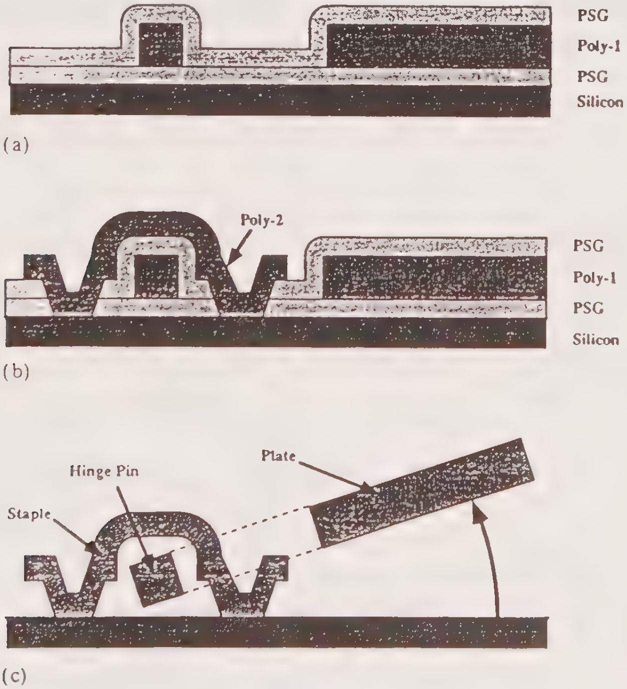
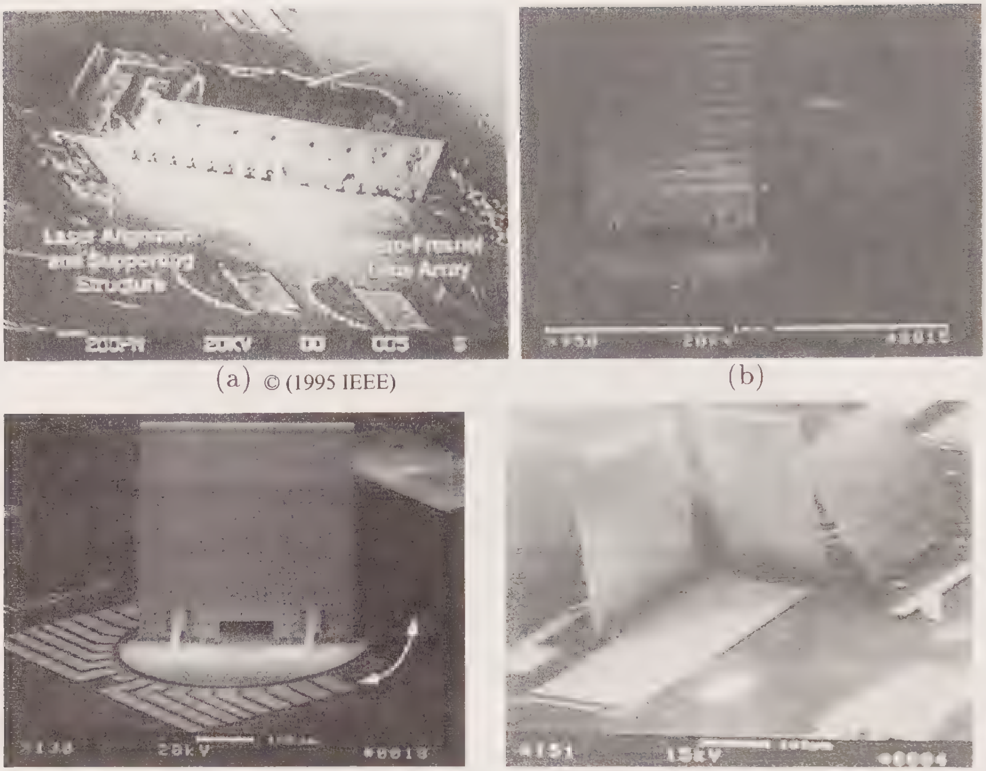

PSG
Poly-1
PSG
Silicon
(a)
Poly-2
PSG
Poly-1
PSG
Silicon
(b)
Plate
Hinge Pin
Staple
(c)

Figure 8.49: Fabrication process of polysilicon hinged structure.
© (Elsevier Science)



(a) © (1995 IEEE) (b)

(c) © (1996 IEEE) (d) © (Elsevier Science)

Figure 8.50: Example of optical applications of polysilicon hinged structures.

Some of the optical applications of polysilicon hinged structures are shown in Fig. 8.50. Figure 8.50(a) is a micro Fresnel lens array coupled with an array of vertical cavity laser diodes developed by Lin et al.[108] and Lee et al.[109]. In another experiment they observed collimation of the light from a single-mode fiber ($\theta = 5.0°$, $\lambda = 1.3$ μm) to 0.33° by the Fresnel lens[110]. They also developed gratings sitting vertically on a rotational stage, shown in Fig. 8.50(b), which were used to split a collimated light in a free space[111]. The angle of the first-order diffraction, at 11~13°, was controlled by the stage rotation. They also used the plate as an etalon, which is a tunable-wavelength filter, as shown in Fig. 8.50(c), and obtained a wide tuning range of 45 nm for a rotation angle of 70°[112]. Figure. 8.50(d) shows a corner cube reflector consisting of three reflecting mirrors by Gunawan et al.[113]. The bottom mirror is actuated by electrostatic attraction of 8 V to modulate the reflected light for an optical communication link in a free space.

These three-dimensional structures can be integrated with microactuators for precision alignment and scanning of the light. For example, Kiang

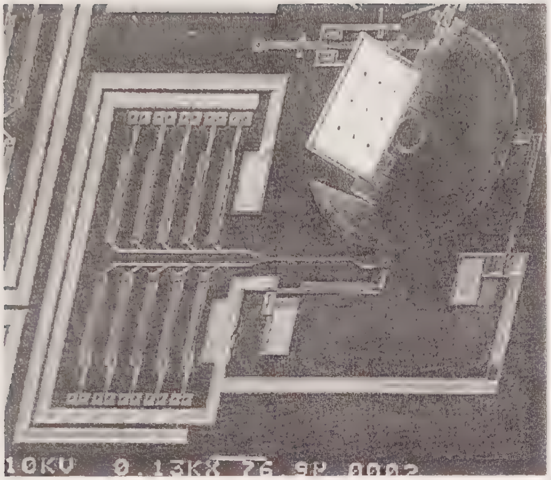


Figure 8.51: Vertical mirror on a step rotor connected to the thermal actuator arrays. © (1996 IEEE)

et al.[74, 75] and Daneman et al.[114] combined a mirror with a comb-drive actuator to construct a scanning device, which was shown in Fig. 8.40. The fiber optic switches and micro-XYZ stages for an MOB developed by Lee et al.[47] and Lin et al.[50] were integrated with scratch drive actuators[49]. Comtois et al. combined their thermal actuator arrays with a polysilicon corner-cube reflector[115] and with a rotating mirror sitting on a stepping rotor[116], as shown in Fig. 8.51.

Some of the three-dimensional structures have been built with microactuators integrated on the same substrate. Daneman et al.[114] used a linear vibromotor to drive a microreflector with an angular travel range of over 90° and a translational travel range of 60 μm. Akiyama et al.[117] proposed to use their scratch drive actuators[49] to build three-dimensional structures assisted by reshaping technology[106].

8.2.5 Microlenses and focusing mirrors

A microlens is one of the fundamental elements in parallel optical processing, such as visual pattern recognition and parallel optical interconnection. Besides the thin-film Fresnel lens arrays, which has been shown in Fig. 8.50(a), various types of microlens arrays have been developed since Oikawa et al.[118, 120] proposed a distributed-index microlens in the early 1980s.

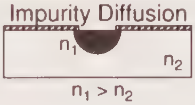
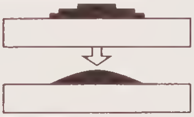
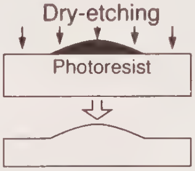
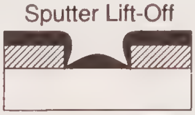
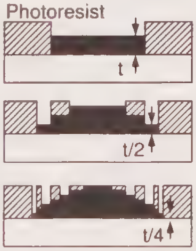
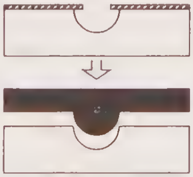
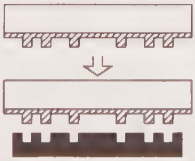
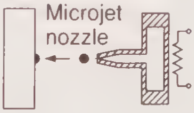
Table 8.5 summarizes the fabrication processes of typical planar lenses. (a) The most fundamental method is monomer exchange in plastic plates[118, 119], ion exchange in optical glasses[127], and electromigration techniques[120].

All of them define two-dimensional refractive index profiles in substrates. (b) Microlenses of large numerical aperture can be formed by a single or multistep etching associated with the mass transport effect at high temperature. Swenson et al.[121] etched a GaP surface in a 15-step pattern and fabricated lenses of 300 μm diameter by smoothing the surfaces at temperatures higher than 1000°C. Liau[122] demonstrated microlens fabrication by the same technique but with one-step etching. (c) Sankur et al.[123] patterned and reflowed photoresist islands on substrates at temperatures of 170~200°C. The photoresist pattern can be used as a microlens as well as a mask pattern to transfer a semispherical profile into a substrate by dry etching. These authors controlled the reactive gas mixture to obtain the optimal etching rates of the photoresist mask material and solid materials (silica, fused silica, silicon, and germanium). (d) Shirai and Serikawa[124] used the sputtering lift-off process to pattern SiO_2 microlens arrays (10×10, $\phi = 800$ μm) of long focal lengths (1~10 cm) on glass substrates. (e) Microlenses of quasi-smooth surfaces or diffractive optical elements can be obtained by constructing multiphase-level structures through the repetition of film deposition. Pawlowski et al.[125] applied 32-level photolithographic steps to develop a binary pattern of sputtered silicon oxide and obtained diffraction lenses of high efficiency (97%). (f) Köehler et al.[126] developed isotropic plasmaless etching of silicon in a gaseous environment of BrF_3 with xenon and obtained a smooth etched surface. With circular holes in the SiO_2 mask, they generated spherical depressions and made replicas of plastic molding which can be used as microlenses. (g) Karthe et al.[127] used a metal master, which was transferred from photoresist patterns by electrodeposition of nickel, to emboss refractive and diffractive lenses onto thermoplastic substrates. (h) A semispherical surface can be obtained by the technique of ink-jet printers. MacFarlane et al.[128] used a molten polymer jet and received the droplets on a soda-lime glass substrate with an XYZ micropositioner. Microlens arrays of diameters ranging between 70 and 150 μm and focal lengths of from 20 to 70 μm were patterned.

Cylindrical lenses can be patterned by anisotropic etching, which Sawada et al.[92] have performed in their optical microencoders: they integrated microlenses with laser diodes and photodetectors on a GaAs substrate, which is seen in Fig. 8.63(a) in the following section.

Micromachined diaphragm devices can be used as a dynamic focusing mirror. Figure 8.52(a) illustrates the concept of the silicon diaphragm focusing mirror developed by Hisanaga et al.[129]. They designed a diaphragm of a three-dimensional thickness profile and used it as an ideal parabolic focusing mirror whose focal length was changed through electrostatic attraction force. The mirror profile was transferred by dry etching from a three-dimensionally shaped photoresist, which had been exposed through a photomask of dither patterns. The diaphragm, ~9 μm thick at the center and 10 mm in diameter,

Table 8.5: Fabrication process of planar microlenses.

Microlens		Process	References
(a)	Distributed Index	 Impurity Diffusion n_1 n_2 $n_1 > n_2$	[118] [119] [120]
(b)	Mass-Transfer		[121] [122]
(c)	Pattern Transfer	 Dry-etching Photoresist	[123]
(d)	Lift Off	 Sputter Lift-Off	[124]
(e)	Multiphase-Level Structures (Binary Optics)	 Photoresist t t/2 t/4	[125]
(f)	Isotropic Etching & Molding		[126]
(g)	Embossing		[127]
(h)	Microjet Fabrication	 Microjet nozzle	[128]

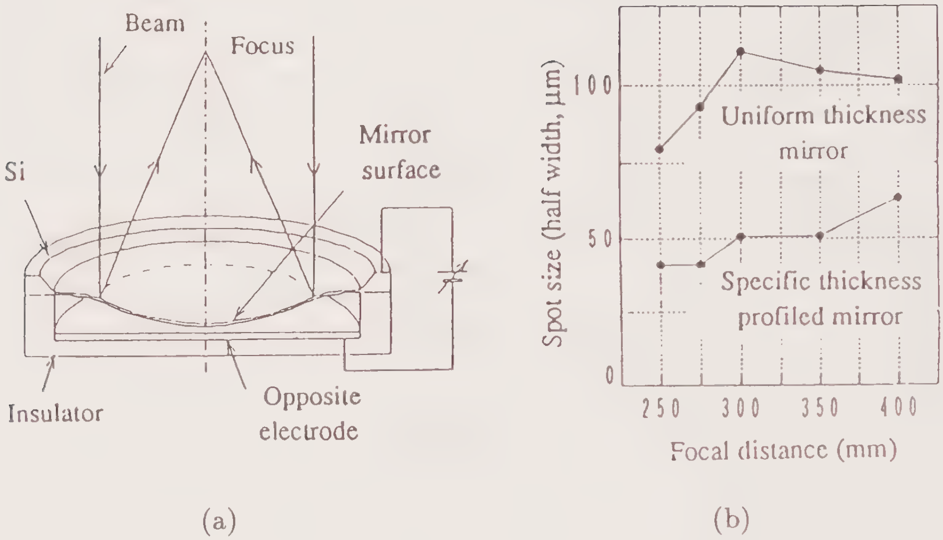


Figure 8.52: Silicon dynamic focusing diaphragm mirror.
© (1993 IEEE)

was dented by 100 μm at the applied voltage of 750 V to provide a focusing spot of 45 μm, which was approximately half that of a uniform diaphragm. The focal length was controlled in the range of 250 to 400 mm, holding the spot size nearly constantly as shown in Fig. 8.52(b).

Figure 8.53 shows a microfocusing mirror based on a piezoelectrically deformable Pb(Zr-Ti)-SiN-SiO_2 diaphragm developed by Goto et al.[130]. A thin film of $Pb_{0.5}Zr_{0.27}Ti_{0.23}$ was deposited by radio frequency(RF)-magnetron sputtering of multiple targets. The piezoelectric material was polarized at 700°C by using electron beam (EB) evaporated Pt-Ti layers as the top and bottom electrodes. The diaphragm was released by KOH etching of the silicon substrate. Goto et al. proposed a concept for a microfocusing device, which consisted of a diaphragm actuator 1.3 mm in diameter and a micro-Fresnel lens. Their calculation showed that a buckling of 4 μm could shift the focal point more than 100 mm. The PZT sputtering and etching conditions are described in Ref. [130].

8.2.6 Optical elements made by the LIGA process

The LIGA process usually utilizes synchrotron orbiter radiation(SOR) X-ray lithography, electroplating, and plastic molding technologies and allows the creation of thick microstructures with submicron accuracy[100, 101, 102]. A thick resist layer, usually a few hundred micrometers and up to a few centimeters in thickness, is exposed through a mask, to an X-ray from an

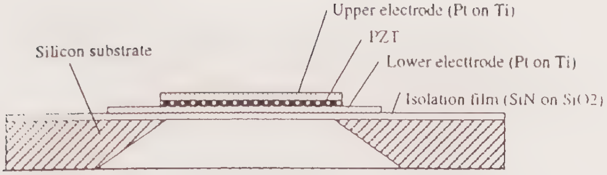

Upper electrode (Pt on Ti)
PZT
Silicon substrate
Lower electrode (Pt on Ti)
Isolation film (SiN on SiO2)

Figure 8.53: PZT thin film actuator as a dynamic focusing mirror.

SOR source. The X-ray runs so straight that fine patterns having exact vertical walls are obtained after development. A conductive substrate is used. Fine metal structures which have shapes complementary to resist patterns are formed on the substrate by electroplating. After removing the resist, one can use metallic structures on the substrate to produce many polymer structures by injection molding or hot embossing. It is possible to fabricate the metallic replica of the polymer structure by electroplating. It is also possible to transfer the pattern into ceramics. The advantages of the LIGA process when it is applied to optics are (1) side walls are smooth and vertical and can be defined with submicron accuracy, (2) transparent materials can be micromachined, and (3) it allows inexpensive mass production of high-precision devices.

The first advantage depends mainly on the lithography step. Since the SOR source is not easy to access, deep ultraviolet(UV) lithography and photosensitive polyimide are sometimes used; these alternatives have limitations in resist thickness, the aspect ratio of structures and wall smoothness. The choice depends on the application requirements. Polymers tend to be transparent and lead to the second advantage. Waveguides can be made. A simultaneous microspectrometer composed of a polymer slab waveguide and a self-focusing reflection grating has been fabricated[100]. The spectrometer, shown in Fig. 8.54, is commercialized as a detector in chemical analysis systems. The molding process, including hot embossing, is known to be a very cost-effective process. Numerous compact disks are made by molding. Three-dimensional microstructures can be obtain by using LIGA-processed masters. Even though making such masters is more expensive than the usual micromachining process, the final products cost much less with the molding process. An optical fiber coupler to an integrated optical chip was made by molding in a LIGA mold[100]. The coupler is shown in Fig. 8.55.

8.2.7 Display

Micromachined displays have attracted much attention recently. A DMD (digital micromirror device) introduced by Texas Instruments has demonstrated excellent performance[5]. It is composed of many movable micromir-

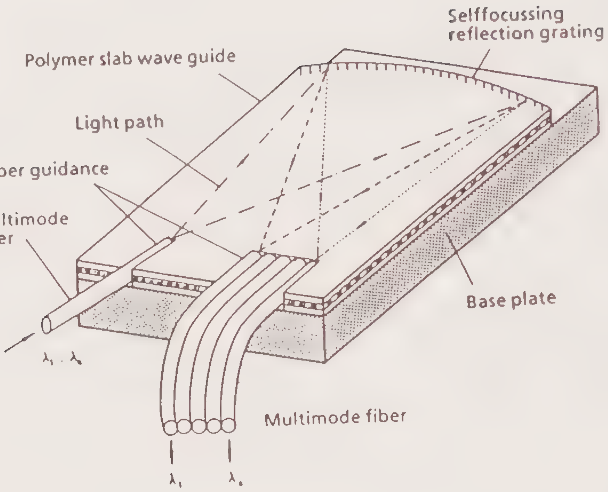

Selffocussing reflection grating
Polymer slab wave guide
Light path
Fiber guidance
Multimode fiber
Base plate
$\lambda_1 \ldots \lambda_n$
Multimode fiber
λ_1
λ_n

Figure 8.54: Schematic representation of a simultaneous microspectrometer by the LIGA process.

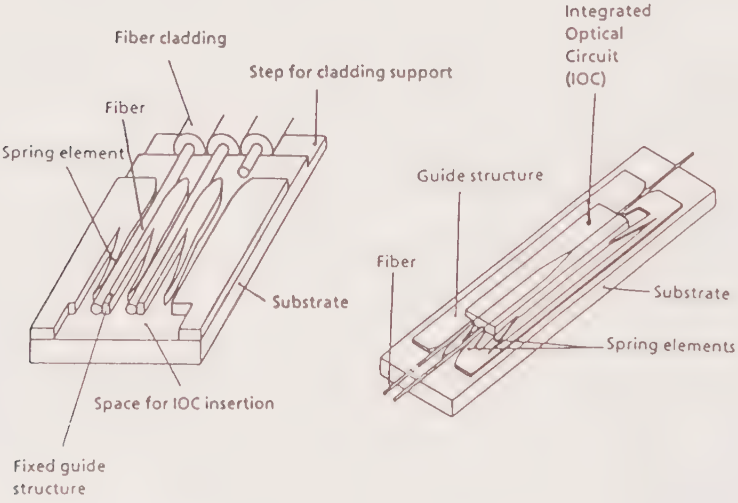

Fiber cladding
Step for cladding support
Fiber
Spring element
Substrate
Space for IOC insertion
Fixed guide structure
Integrated Optical Circuit (IOC)
Guide structure
Fiber
Substrate
Spring elements

Figure 8.55: Fiber-chip coupling elements by the LIGA process.

© (Springer-Verlag)

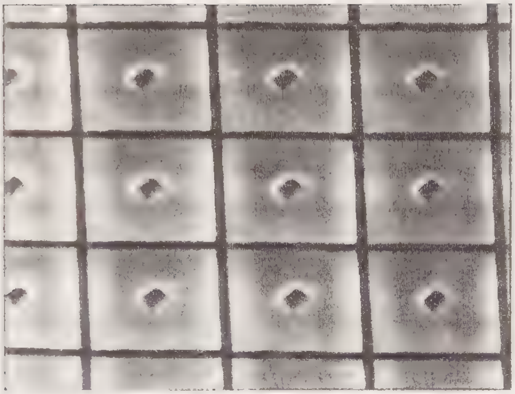


Figure 8.56: A portion of the tiling pixel array of a digital micromirror device.

rors integrated on CMOS circuits. Both the device and the fabrication technology will have wide varieties of applications[87]. DMDs are applied to projection displays and laser printers and can be used for image processing and modulation. The technology enables integration of mechanical elements with control circuits; this leads to smart microsystems.

Each micromirror of a DMD is supported by elastic hinges that allow torsion or linear deflection of the mirror by electrostatic force. Torsional mirror arrays are used for display. The mirror has binary states of ±10°. Incident light from an angle of 20° is reflected perpendicular to the DMD at mirrors in a +10° state and projected to the screen. Those pixels which correspond to such mirrors are bright. Light reflected at mirrors in a −10° state is eliminated by an aperture. Eight-bit pulse width modulation(PWM) of the mirror state produces gray or color images. Three DMDs or one DMD with a rotational color filter in front of the illumination is used for color images.

An SEM picture of the mirror array is shown in Fig. 8.56. The side of each mirror is 16 μm. DMDs for the NTSC (National Television System Comittiee) have 768×576 mirrors, and those for VGA (video graphic array), SVGA (super video graphic array) and HDTV (high definition television) have 1280×1024 or even 2048×1152 mirrors. Figure 8.57 shows the structure of the mirror. There are three layers for mechanical parts; note that these layers are built on top of CMOS circuits. The lowest layer has electrodes to supply potentials to addressing electrodes and the mirror. Addressing electrodes and a movable base supported by thin torsion hinges are in the second layer. Both ends of the base touch the first layer when electrostatic actuation is performed. This determines the tilting angle of the mirror, which is fixed to the base through a post and conceals all the structures in

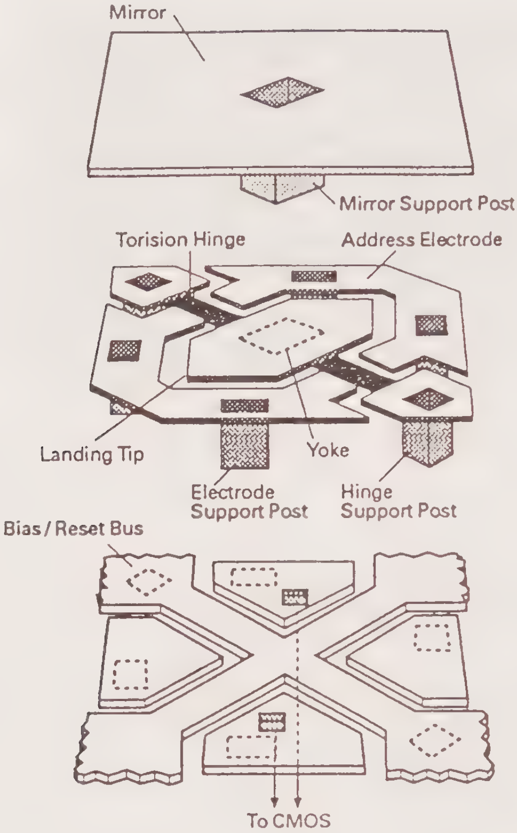

Mirror
Mirror Support Post
Torision Hinge
Address Electrode
Landing Tip
Yoke
Electrode Support Post
Hinge Support Post
Bias / Reset Bus
To CMOS

Figure 8.57: Exploded view of the DMD.

the first and second layers. This hidden hinge structure improves the area efficiency and contrast ratio. All structures are made with Al, and polymer sacrificial layers are used. The process temperature is low enough to ensure compatibility with the CMOS circuit built in the substrate.

The response time of the mirror is ~10μs. The speed is fast enough for a 768×576 pixel NTSC video display. The lifetime of the mirrors was confirmed. Packaging of the device is important to avoid damage. The high intensity illumination, which can excite hot carriers through heat and photons in SRAM circuits, has proved to cause no problem.

The DMD is a good example of the advantages of MEMS. The miniaturization of the mirror enables a high speed response. Multiple use of simple mirrors gives us images on a screen. Microelectronics integrated with mechanical parts can control many mirrors individually and quickly.

Vdovin and Middelhoek[88] deformed a flat, thin membrane locally and used it as a display. The device consists of a freely suspended SiN-Al re-

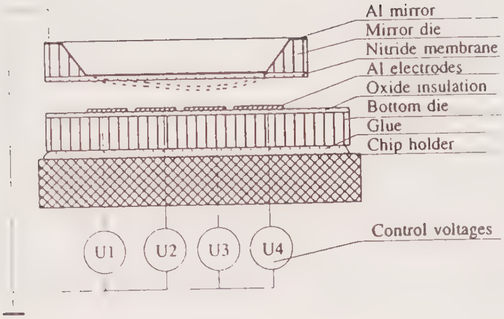


Figure 8.58: Schematic view of deformable mirror display device.
© (1995 IEEE)

flecting membrane and an array of driving electrodes. Figure 8.58 shows the cross-section of the device schematically. A nitride membrane (10 mm × 10 mm ×0.5 μm) was etched anisotropically by KOH and covered with an Al layer 0.2 μm thick on the etched side. The chip with the membrane is fixed 25~100 μm above the driving electrodes on another chip. The optical figure of the membrane can be locally controlled by the voltages at the electrodes. The intensity distribution of the reflected white beam follows approximately the driving voltage distribution. As shown in Fig. 8.59, the intensity distribution on the Z_1 plane can be projected into a remote plane Z_2 with the projection lens L. A contrast ratio of 3:1 and resolution of approximately 100 by 100 pixels with a response time of ~1 ms have been demonstrated. An example of such an image is shown in Fig. 8.60. It is concluded that the device could make high-resolution halftone images with comparatively low contrast.

Huang et al.[89] fabricated a 40×40 cantilever array which could be driven piezoelectrically. The pixel size is 100 μm×100 μm. ZnO was used as the piezoelectric film. They obtained a 1.9 μm deflection with 10 Vp-p pulse at 10 Hz; however no optical image was given in the paper.

8.2.8 Micro-optical disk

MEMS are well suited to data storage devices because a data bit has no mass or size in principle. The size of data bits is limited only by read-write technology and thermal stability of the mark. A mark size of a few nanometers is possible with scanning probe microscopes such as a scanning tunneling microscope(STM) or atomic force microscope(AFM), although the read/write speed is still too slow for practical use.

Recentry, microactuators have been used in magnetic disk drives[97, 98].

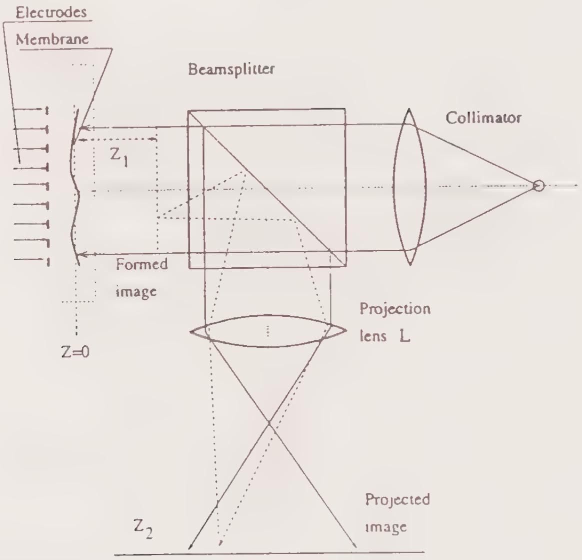

Electrodes
Membrane
Beamsplitter
Collimator
Z_1
Formed
image
Z=0
Projection
lens L
Projected
image
Z_2

Figure 8.59: Schematic of deformable mirror display.
© (1995 IEEE)

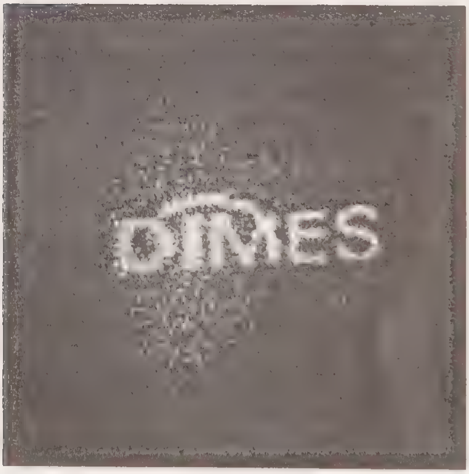


Figure 8.60: Image obtained in experiment by the DMD.
© (1995 IEEE)

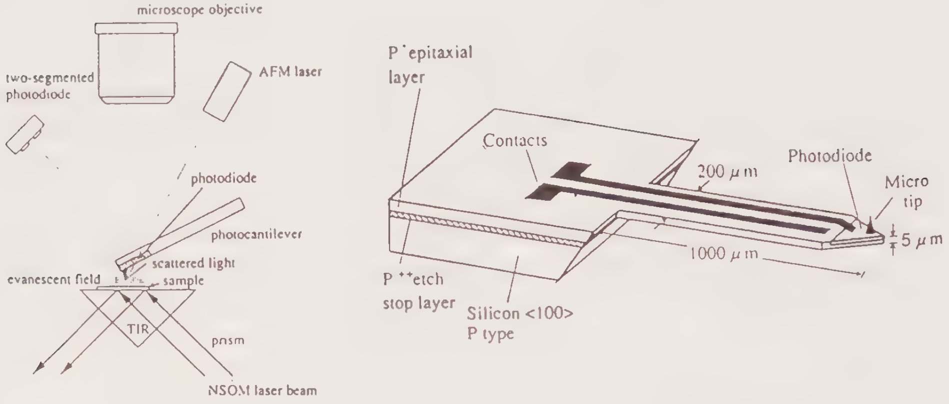
There is a trial under way to apply a micromotor to an optical disk device. Mourlas et al.[99] fabricated electrostatic micromotors on which small pits with sizes of 2~20 μm were patterned. The diameter of the motor was 500 μm. It rotated at 6300 rpm with a driving voltage of 100 V. They succeeded in reading out the bit with sizes over 5 μm and confirmed the feasibility of micro-optical disk drives. Further development to reduce spot sizes and to achieve precise motor speed and position control must be done. The application opportunities for MEMS in optical disks are not limited to rotating the disk but include positioning the head, focusing a beam, and fabricating an integrated head. It is also proposed to use near-field scanning optical microscope(NSOM) technology for very high storage density (see the next section).

8.2.9 Near-field scanning optical microscope

There are wide varieties of scanning probe microscopes (SPM) for observing surface properties with atomic-scale resolution. Near-field scanning optical microscopes are one of the SPM family. An optical fiber with a protruding sharp core is usually used as the probe. Akamine et al.[90, 91] developed a microcantilever 100 μm long, 200 μm wide and 5 μm thick on which a photodetector is integrated. The schematic representation of the cantilever is shown in Fig. 8.61(a). The cantilever detects evanescent light from an illuminated sample as shown in Fig. 8.61(b). It can be used at the same time as the probe of an atomic force microscope. They have obtained NSOM and AFM images of a sample with a resolution of 1 nm order. Figure 8.62 shows the image of a CD grating and dust particles on it. In an expanded image, the details of the particles were resolved with sizes of 20 nm. They confirmed that the SNOM image and the AFM image corresponded very well.

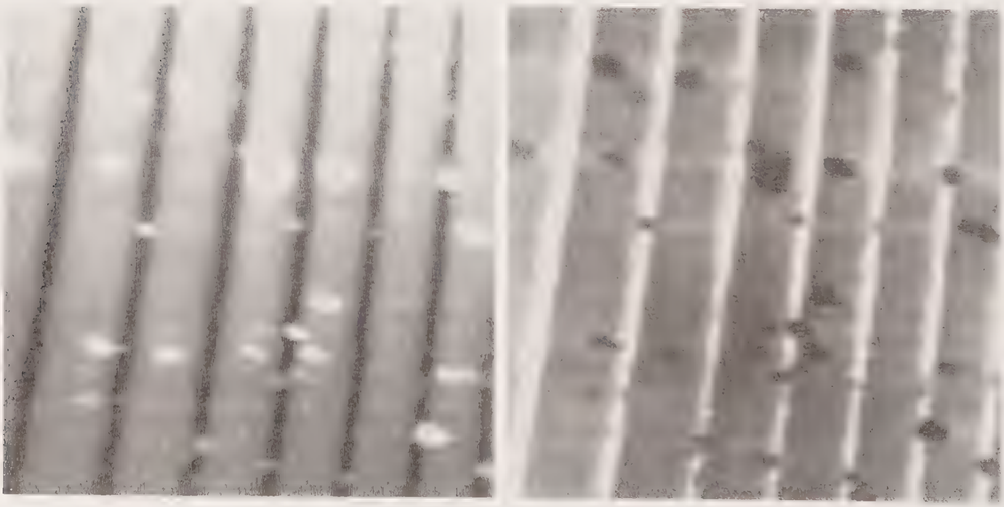
8.2.10 Optical microencoder and displacement sensor

Sawada et al.[92] and Sawada[93] developed an integrated optical microencoder. They integrated a U-shaped laser diode with etched mirrors, optical waveguides made of polyimide, and two photodiodes as shown in Fig 8.63. The ends of waveguides are patterned in a convex shape and work as lenses. The output beam has a diverging angle of $1° \times 8°$. The air gap between the encoder and the grating is less than 1 mm. A small step is etched in one of these ends to give a 90° phase shift to the beam emitted from the end. The direction of movement was detected by using two beams with a 90° phase difference. Please note that this sensor has all the functions of a total system. The size is less than 1 mm×2 mm. These authors claimed a theoretical resolution of 0.01 μm with a 1 μm-pitch grating. Because of its miniature size and the fabrication process, it may be possible to integrate the encoder with microactuators, which will result in a micropositioner with very high


microscope objective
two-segmented photodiode
AFM laser
photodiode
photocantilever
scattered light
evanescent field
sample
TIR
prism
NSOM laser beam
P⁻epitaxial layer
Contacts
200 μm
Photodiode
Micro tip
5 μm
1000 μm
P⁺⁺etch stop layer
Silicon <100> P type

(a) (b)

Figure 8.61: Concept of near-field scanning optical microscope by microphoto cantilever. © (1995 IEEE)



(a) (b)

Figure 8.62: CD disk grating images obtained by (a) AFM and (b) NSOM. Scanned area is 10 μm×10 μm. © (1995 IEEE)

accuracy. Multiple encoders can be attached to a macroscopic machine to control its linear and angular motion precisely. If this idea is applied to a stepper for photolithography, interferometers will no longer be necessary. The integration of the encoder and a silicon IC chip for signal processing is possible in principle; it will result in a complete measurement system on a single chip.

Miyajima et al.[94] developed a microencoder using a surface-emitting laser. The surface-emitting laser consists of $In_{0.12}Ga_{0.88}$ As strained quantum wells sandwiched by front and rear mirror stacks. The output power is 0.5 mW with 35 mA. The wavelength is 980 nm. A photodiode to detect a reflected beam from a scale was fabricated on a silicon chip ($1.8\times1.4\times0.45$ mm^3). The surface-emitting laser was attached on the chip by eutectic bonding. A quasi-sinusoidal signal was obtained with a scale of 20 μm in pitch. These authors also fabricated a microlens which could be integrated on the laser. A positive photoresist layer was patterned into circular shapes and baked at a high temperature. Thermal reflow at the temperature resulted in a spherical profile of the surface. The aperture size and the focal length were designed to be 30$\sim$50 μm and 300$\sim$500 μm, respectively. They expect a better resolution of 0.2 μm with the lens.

Zappe et al.[95, 96] fabricated an integrated optical system to measure displacement based on the Michelson interferometer. It is another type of optical displacement sensor. A distributed Bragg reflector laser, optical waveguides and a photodetector were monolithically fabricated on a GaAs substrate, as schematically shown in Fig. 8.64. The wavelength of the laser was 820 nm. Maximum optical output exceeded 5 mW. A graded index (GRIN) lens was aligned at the output. Zappe et al. observed 49 fringes over 20 μm displacement of an object placed 130 mm away from the lens.

8.2.11 Waveguide sensors integrated with MEMS structures

Besides the encoders and displacement sensors in the previous section, various kind of sensors have been developed as a microfabricated waveguide structure. The principles of sensing are similar to the mechanisms of waveguide switches and modulators described in Section 8.2.2, i.e., some of them are based on phase modulation by using Mach-Zender[155, 156, 157, 158, 159, 160] or Michelson[95, 96, 161] interferometers, and the others are based on intensity modulation using the coupling of waveguide cantilevers[162, 163], evanescent modulation[164], or a strain-sensitive Bragg grating waveguide[165].

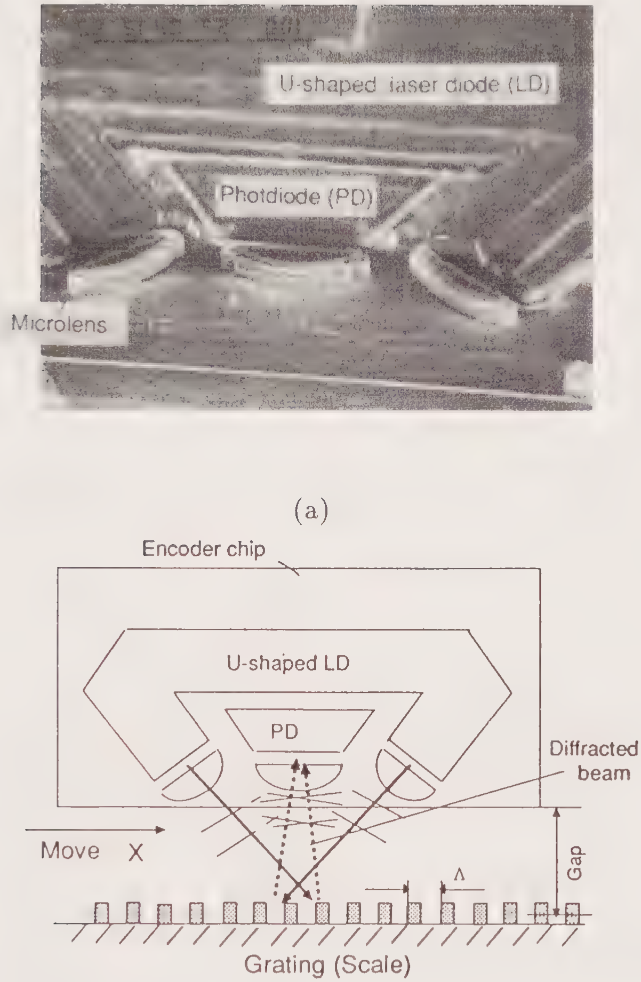
Ohkawa et al.[155] patterned a waveguide on a silicon diaphragm to create a pressure-sensitive Mach-Zender interferometer, as illustrated in Fig. 8.65 (a). The pressure difference deforms the diaphragm to change the refractive index, and thus changes a phase shift of the lightwave propagating in the waveguide on the diaphragm. The guided wave in the other branch is not affected by the pressure because the guide is placed off the diaphragm. These waveguides are merged into the output waveguide to transform the phase difference, i.e., applied pressure into the output light intensity. The rib waveguide (single-mode for the 633-nm wavelength) was made of sputtered glass, and the diaphragm (1 mm long and 220 μm wide) was formed by anisotropic etching in ethylene diamine, pyrocatechol, and water(EDPW) of the p^{++} etch stop technique. Figure 8.65(b) shows the measured output as a function of applied pressure. The half-wave pressure was 0.8×10^5 Pa and the extinction ratio was 10 dB. No diaphragm failure was found in the measured pressure range. Recently Nathan et al.[160] have developed a similar type of pressure sensor with an ARROW waveguide.

By modifying the diaphragm in Fig. 8.65 into a planar cantilever (3 mm long, 5 mm wide and 15 μm thick), Ohkawa et al.[156] have shown that the Mach-Zender interferometer can be used as a flow sensor. Figure 8.66 shows the cross-section of the device and the output measured in relation to the liquid flow rate. They observed a decrease in the light intensity at a flow rate up to 600 cc/min. Flow rates over this value caused the oscillation of the cantilever, disturbing measurement. They also proposed an accelerometer with the same configuration but with a proof mass on the backside of the cantilever[157].

Boiarski et al.[159] used a Mach-Zender interferometer of an ion-exchanged glass waveguide as a biosensor. Figure 8.67 illustrates the device configuration. One of the branches is protected with a buffer layer of glass, and the other is coated with an antibody layer that selectively binds with an analyte molecules. When the sensing waveguide is exposed to analyte (protein immunoglobulin G in their experiment), the effective refractive index on the waveguide surface is changed. They could detect the protein at a level of 1 μg/ml in less than 10 minutes; this provides great convenience in biochemical experiments compared with the conventional methods.

Wu and Frankena[163] patterned a waveguide into cantilever and bridge structures and demonstrated acoustic signal sensing. The device is made of SiO_2-Al_2O_3-SiO_2 layers patterned into a cantilever 100 μm long, 10~30 μm wide and 2.5 μm thick. The modulation mechanism is similar to that of the device developed by Bezzaoui and Voges[58] and Hoffmann et al.[59], which was shown in the previous section. They also mentioned in their paper that the cantilever sensor could be used as a scanning force microscope or a surface profiler which was sensitive enough to detect a displacement of 0.1 Å and a force of 10^{-10} N. Burcham et al.[162] proposed using a waveguide structure similar to an accelerometer.

Young et al.[164] developed a design of microfabricated pressure transducers based on evanescent modulation. The device is composed of the two pieces of silicon substrates: a bottom substrate of the waveguide and a top substrate of a pressure-sensitive diaphragm. The mechanism of sensing, i.e.,


U-shaped laser diode (LD)
Photdiode (PD)
Microlens
Encoder chip
U-shaped LD
PD
Diffracted beam
Move X
Gap
Λ
Grating (Scale)

(a)

(b)

Figure 8.63: (a) SEM view of integrated optical encoders and (b) measurement principle. © (Elsevier Science)

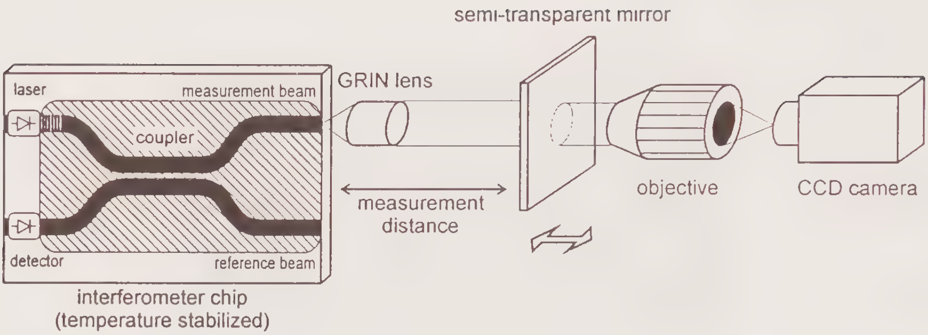
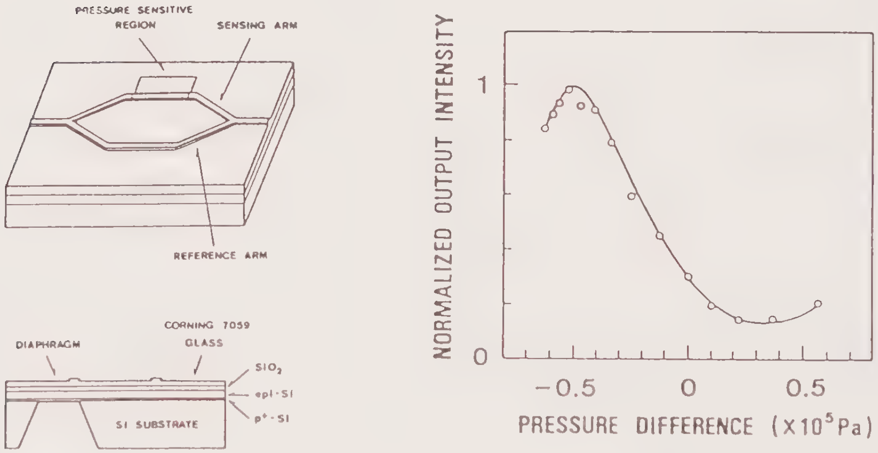

semi-transparent mirror
laser
measurement beam
coupler
detector
reference beam
interferometer chip
(temperature stabilized)
GRIN lens
measurement
distance
objective
CCD camera

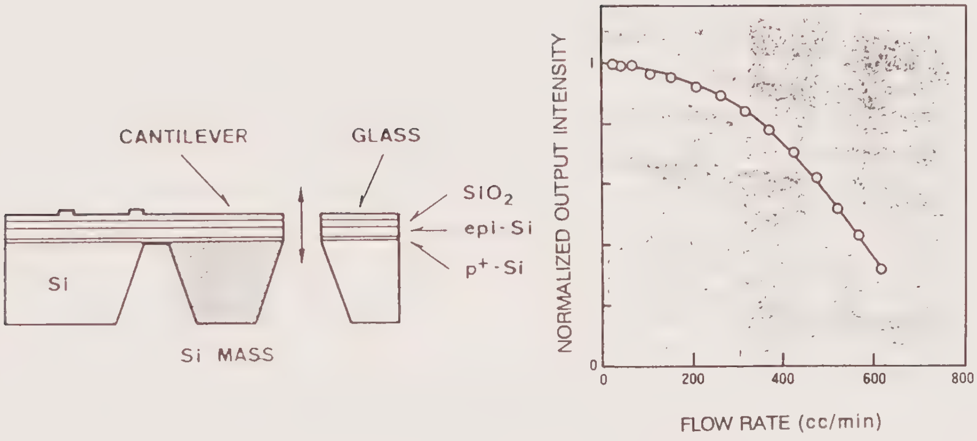
Figure 8.64: Measurement setup to characterize the integrated interferometer for displacement sensing. © (1996 IEEE)



(a)

(b)

Figure 8.65: Integrated optic pressure sensor. (a) schematic view and (b) measured output intensity as a function of applied pressure difference.



(a) (b)

Figure 8.66: Mach-Zender interferometer as a flow sensor. (a) cross-section of the structure and (b) measured output as a function of flow rate.

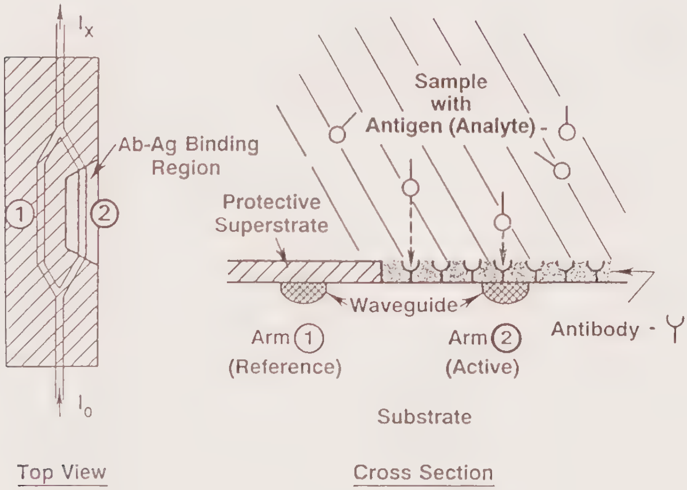

I_x
Ab-Ag Binding Region
①
②
I_0
Sample with Antigen (Analyte) -
Protective Superstrate
Waveguide
Arm ① (Reference)
Arm ② (Active)
Antibody -
Substrate
Top View
Cross Section

Figure 8.67: Mach-Zender interferometer as a biosensor.

evanescent coupling and spoiling, was described in Section 8.2.2. Storgaard-Larsen et al.[165] employed a Bragg grating patterned into a waveguide bridge as the strain-sensing element of an accelerometer.

These types of microfabricated sensing elements can be also combined with optical fibers to create fiber optic accelerometers, hydrophonic sensors, pressure sensors, and electric and magnetic field sensors. Most of the devices' function can be understood by the mechanism of the Fabry-Perot interferometer or coupling of fibers. References for typical examples are listed in Table 8.1.

8.3 Conclusion

The application of micromachine technologies to micro-optical devices has been reviewed. A basic concept making the best use of the advantages of micromachining technologies is proposed. Recent examples of optical microelectromechanical systems are introduced. Most of them are still components or subsystems. Further development is anticipated when total systems, including active devices such as laser diodes, can be easily fabricated and integrated with MEMS devices.

References

[1] R. T. Howe, R. S. Muller, K. J. Gabriel, W. S. N. Trimmer, "Silicon Micromechanics: Sensors and Actuators on a Chip," *IEEE Spectrum*, 29-35 (1991).

[2] K. J. Gabriel, "Engineering Microscopic Machines," *Scientific American* **260** (9),118-121 (1995).

[3] H. Fujita, "Recent Progress in Micromachining and Applications to Microactuators," *Jpn. J. Appl. Phys.* **33**, part 1, no. 12B, 7163-7166 (1994).

[4] K. E. Petersen, "Micromechanical light modulator array fabricated on silicon," *Appl. Phys. Let.* **31** (8), p. 521 (1977).

[5] J. M. Younse, "Mirrors on a chip," *IEEE Spectrum*, p. 27, November (1993).

[6] A. Aratani, P. J. French, P. M. Sarro, D. Poenar, R. F. Wolffenbuttel, S. Middelhoek, "Surface micromachined tunable interferometer array," *Sensors and Actuators* **A 43**, 17-23 (1994).

[7] A. Aratani, P. J. French, P. M. Sarro, D. Poenar, R. F. Wolffenbuttel, S. Middelhoek, "Process and design considerations for surface micromachined beams for a tunable interferometer array in silicon," *Proc. IEEE Micro Electro Mechanical Systems (MEMS 93)*, Fort Lauderdale, Florida, Feb. 7-10 (1993), 230-235.

[8] J. H. Jerman, D. J. Clift, S. R. Mallinson, "A Miniature Fabry-Perot Interferometer with a Corrugated Diaphragm Support," *Sensors and Actuators* **A-29** (2) 151-158 (1991).

[9] M. C. Larson, B. Pezeshiki, J. S. Harris, Jr., "Vertical Coupled-Cavity Micro Interferometer on GaAs with Deformable-Membrane Top Mirror," *IEEE Photo. Tech. Lett.* **7** (4), 382-384 (1995).

[10] M. C. Larson, J. S. Harris, Jr., "Broadly-Tunable Resonant-Cavity Light-Emitting Diode," *IEEE Photo. Tech. Lett.* **7** (11), 1267-1269 (1995).

[11] Y. Uenishi, M. Tsugai, M. Mehregany, "Hybrid-integrated laser-diode micro-external mirror fabricated by (110) silicon micromachining," *Electron. Lett.* **31** (12), 965-966 (1995).

[12] Y. Uenishi, M. Tsugai, M. Mehregany, "Micro-opto-mechanical devices fabricated by anisotropic etching of (110) silicon," *J. Micromech. Microeng.* **5**, 305-312 (1995).

[13] Y. Uenishi, K. Honma, S. Nagaoka, "Tunable laser diode using a nickel micromachined external mirror," *Electron. Lett.*, **32** (13), 1207-1208 (1996).

[14] K. Hjort, J. Söderkvist, J-A,Schweiz, "Gallium arsenide as a mechanical material," *J. Micromech. Microeng.* **4**, 1-13 (1994).

[15] J. Miao, K. Hjort, H.-L. Hartnagel. J.-A. Schweits, D. Ruck, K. Tischert, "Resonant sensors on thin semi-insulating GaAs membranes," *Tech. Digest 8th Int. Conf. on Solid-State Sensors and Actuators (TRANSDUCERS 95)*, Stockholm, Sweden, Jun. 25-29 (1995), **vol. 2**, 604-607.

[16] K. Hjort, J.-A. Schweits, B. Hok, *IEEE MEMS '90*, p. 73 (1990).

[17] Z. L. Zhang, G. A. Porkoab, N. C. MacDonald, *IEEE MEMS '92*, **72** (1992).

[18] Y. Uenishi, H. Tanaka, H. Ukita, "Monolithic integration of microbeam resonators and laser diodes using AlGaAs/GaAs micromachining," *MICRO SYSTEM Technologies 90*, (ed. H. Reichl), VDE-Verlag, pp. 313-322 (1990).

[19] Y. Uenishi, H. Tanaka, H. Ukita, "Characterization of AlGaAs Microstructure Fabricated by AlGaAs/GaAs Micromachining," *IEEE Trans. Electron Device* **41** (10), 1778-1783 (1994).

[20] K. Iga et al., *IEEE Quantum Electron.*, **QE-23**, 1845 (1988).

[21] J. L. Jewell, J. P. Harbison, A. Scherer, Y. H. Lee, L. T. Florez, "Vertical-cavity surface-emitting lasers: Design, growth, fabrication, characterization," *IEEE J. Quantum Electron.* **27**, 1332-1346 (1991).

[22] J. S. Harris, Jr., M. C. Larson, A. R. Massengale, "Broad-Range Continuous Wavelength Tuning in Microelectromechanical Vertical-Cavity Surface-Emitting Lasers," *Digest IEEE/LEOS 1996 Summer Topical Meetings*, Keystone, Colorado, Aug. 5-9 (1996), 31-32.

[23] E. Yamamoto, S. Hashimoto, M. Ito, I. Komazaki, K. Yanagisawa, "Optical Tactile Sensor Using Surface-Emitting Laser," *Proc. IEEE Micro Electro Mechanical Systems (MEMS 95)*, Amsterdam, the Netherlands, Jan. 29-Feb. 2 (1995), 227-231.

[24] M. Ito, E. Yamamoto, S.H ashimoto, I. Komazaki, H. Miyajima, S. Shinohara, K. Yanagisawa, "Compound-Cavity Micro-Interferometer using VCSEL and Piezoelectric Actuator," *Digest IEEE/LEOS 1996 Summer Topical Meetings*, Keystone, Colorado, Aug. 5-9 (1996), 51-52.

[25] M. S. Wu, E. C. Vail, G. S. Li, W. Yuen, C.J.Chang-Hasnain, "Widely and Continuously Tunable Micromachined Resonant Cavity Detector with Wavelength Tracking," *IEEE Photon. Tech. Lett.* **8** (1), 98-100 (1996).

[26] C. J. Chang-Hasnain. E. C. Vail, M. S. Wu, "Widely-Tunable Micro-Mechanical Vertical Cavity Lasers and Detectors," *Digest IEEE/LEOS 1996 Summer Topical Meetings*, Keystone, Colorado, Aug. 5-9 (1996), 43-44.

[27] J. A. Walker, K. W. Goossen and S. C. Arney, "Fabrication of a Mechanical Anti-Reflection Switch for Fiber-to-the-Home Applications," *IEEE/ASME J. of Microelectromechanical Syst.* **5** (1), 45-51 (1996).

[28] K. W. Goossen, J. A. Walker, S. C. Arney, "Silicon Modulator Based on Mechanically-Active Anti-Reflection Layer with 1 Mbit/sec Capability for ibr-in-the-Loop Applications," *IEEE Photon. Tech. Lett.* **6** (9), 1119-1121, (1994).

[29] J. A. Walker, K. W. Goossen, P. P. Iannone, W. M. MacDonald, R. Ruel, R. Boie, S. C. Arney and N. J. Frigo, "3.5 Mbit/sec. MARS Modulator System Performance for Fiber-to-the-Home Applications," *Solid-State Sensor and Actuator Workshop*, Hilton Head Island, South Carolina, June 3-6 (1996), pp. 35-36.

[30] J. A. Walker, K. W. Goossen, S. C. Arney, "Mechanical Anti-Reflection Switch (MARS) Device for Fiber-in-the-Loop Applications," *Digest IEEE/LEOS 1996 Summer Topical Meetings*, Keystone, Colorado, Aug. 5-9 (1996), 59-60.

[31] O. Anthamatten, R. K. Battig, B. Valk, P. Vogel. C. Marxer, M. Grétillat, N. F. de Rooij, "Packaging of a reflective optical duplexer based on silicon micromechanics," *Digest IEEE/LEOS 1996 Summer Topical Meetings*, Keystone, Colorado, Aug. 5-9 (1996), 61-62.

[32] C. Marxer, M.Grétillat, V. P. Jaecklin, R. Baettig, O. Anthamatten, P. Vogel and N. F. de Rooij, "MHz Opto-Mechanical Modulator," *The 8th Int. Conf. on Solid-State Sensors and Actuators (TRANSDUCERS 95)*, Stockholm, Sweden, Jun. 25-29 (1995), **vol. 1**, 289-292.

[33] W. R. Wiszniewski, R. E. Collins, B. A. Pailthorpe, "Mechanical light modulator fabricated on a silicon chip using SIMOX technology," *Sensors and Actuators* **A 43**, 170-174 (1994).

[34] N. C. MacDonald, A. Jazairy, "Single crystal silicon: application to micro-opto-electromechanical devices," *SPIE Micro-Optics/Micromechanics and Laser Scanning and Shaping*, vol. 2383, 125-135 (1995).

[35] V. P. Jaecklin, C. Linder, N. F. de Rooij, J.-M. Moret, R. Vuilleumier, "Line-addressable torsional micromirrors for light modulator arrays," *Sensors and Actuators* **A 41-42**, 324-329 (1994).

[36] Y.-A. Peter, H.-P. Herzig, E. Rochat, R. Dändliker, C. Marxer, N. F. de Rooij, "Q-switched fiber laser using a torsional micro-mirror," *Digest IEEE/LEOS 1996 Summer Topical Meetings*, Keystone, Colorado, Aug. 5-9 (1996), 67-68.

[37] O. Solgaard, F. S. Sandejas, D. M. Bloom, "Deformable grating optical modulator," *Optics Letters* **17** (9), 688-690 (1992).

[38] D. E. Sene, J. W. Grantham, V. M. Bright, J. H. Comtois, "Development and characterization of micro-mechanical gratings for optical modulation," *Proc. IEEE Micro Electro Mechanical Systems (MEMS 96)*, San Diego, California, USA, Feb. 11-15 (1996), 222-227.

[39] R. B. Apte, F. S. C. Sandejas, W. C. Banyai and D. M. Bloom, "Deformable Grating Light Valves for High Resolution Displays," *Solid-State Sensor and Actuator Workshop*, Hilton Head Island, South Carolina, June 13-16 (1994), 1-6.

[40] K. Hogari, T. Matsumoto, "Electrostatically Driven Fiber-Optic Micromechanical On/Off Switch and Its Application to Subscriber Transmission Systems," *IEEE Journal of Light Wave Tech.* **8** (5), 722-727 (1990).

[41] R. Jebens, W. Trimmer and J. Walker,"Microactuators for aligning optical fibers," *Sensors and Actuators* **20**, 65-73 (1989).

[42] S. Nagasawa, H. Furukawa, T. Satake, N. Kashima, "New type of optical switch with a plastic-molded ferrule," *Trans. Inst. Electron. Inf. Commun. Eng.* **E70** (8), 696-698 (1987).

[43] S. Nagaoka, "Micro-Magnetic Alloy Tubes for Switching and Splicing Single-Mode Fibers," *Proc. 4th IEEE Workshop on Microelectro mechanical Systems (MEMS 91)*, Nara, Japan, Jan. 30-Feb. 2 (1991), 86-91.

[44] S. Nagaoka, "Optomechanical Switches for Fiber-Optic Communication Systems," *Proc. SPIE Int. Conf. on Optical Fabrication and Testing*, vol. 2576, 383-394, (1995).

[45] L. A. Field, D. L. Burriesci, P. R. Robrish, R.C. Ruby, "Micromachined 1×2 optical-fiber switch," *Sensors and Actuators* **A 53**, 311-315 (1996).

[46] M. F. Dautartas, A. M. Benzoni, Y. C. Chen, G. E. Blonder, B. H. Johnson, C. R. Paola, E. Rice, Y.-H. Wong, "A Silicon-Based Moving-Mirror Optical Switch," *Journal of Lightwave Tech.* **10** (8), 1078-1085 (1992).

[47] S. S. Lee, L. Y. Lin, M. C. Wu, "Surface-micromachined free-space fiber-optic switches," *Elec. Lett.* **31** (17), 1481-1482 (1995).

[48] K. S. J. Pister, M. W. Judy, S. R. Burgett, R. S. Fearing, "Microfabricated hinges," *Sensors and Actuators* **A 33**, 249-256 (1992).

[49] T. Akiyama, K. Shono, "Controlled stepwise motion in polysilicon microstructures," *IEEE J. Microelectromechanical Systems* **2** (2), 106-110 (1993).

[50] L. Y. Lin, J. L. Shen, S. S. Lee, M. C. Wu, "Surface-Micromachined Micro-XYZ Stages for Free-Space Micro-Optical Bench," *Digest IEEE/LEOS 1995 Summer Topical Meetings*, Keystone, Colorado, Aug. 5-9 (1996), 37-38.

[51] J. Mohr, M. Kohl, W. Menz, "Micro Optical Switching by Electrostatic Linear Actuators with Large Displacements," *Proc. 7th Int. Conf. on Solid-State Sensors and Actuators (TRANSDUCERS 93)*, Yokohama, Japan, Jun. 7-10 (1993), 120-123.

[52] R. M. Boysel, T. G. McDonald, G. A. Magel, G. C. Smith, J. L. Leonard, "Integration of Deformable Mirror Devices with optical fibers and waveguides," *SPIE Integrated Optics and Microstructures*, vol. 1793, 34-39 (1992).

[53] H. Toshiyoshi, H. Fujita, "Electrostatic Micro Torsion Mirrors for an Optical Switch Matrix," *IEEE J. Microelectromechanical Sys.* **5** (4), 231-237 (1996).

[54] T. Kanai, A. Nagayama, S. Inagaki, K. Sasakura, "Automated Optical Main Distributing-Frame System," *OFC/IOOC Tech Digest*, 4 TuO5, 74-75 (1993).

[55] J. L. Jackel, J. J. Johnson, W. J. Tomlinson, "Bistable optical switching using electrochemically generated bubbles," *Optics Letters* **15** (24), 1470-1472 (1990).

[56] M. Sato, "Electrocapillarity optical switch," *IEICE Trans. Commun.* **E77-B** (2), 197-203 (1994).

[57] K. Hogari, T. Matsumoto, "Electrostatically driven micromechanical 2×2 optical switch," *Appl. Optics.* **30** (10), 1253-1257 (1991).

[58] H. Bezzaoui, E. Voges, "Integrated optics combined with micromechanics on silicon," *Sensors and Actuators* **A 29**, 212-223 (1991).

[59] M. Hoffmann, H. Bezzaoui, E. Voges, "Micromechanical cantilever resonators with integrated optical interrogation," *Sensors and Actuators* **A 44**, 71-75 (1994).

[60] T. T. H. Eng, S. Y. S. Sin, S. C. Kan, G. K. L. Wong, "Micromechanical Optical Switching with Voltage Control Using SOI Movable Integrated Optical Waveguides," *IEEE Photon. Tech. Lett.* **7** (11), 1297-1299 (1995).

[61] E. Ollier, P. Labeye, F. Revol, "Micro-opto mechanical switch integrated on silicon," *Electron. Lett.* **31** (23), 2003-2005 (1995).

[62] E. Ollier, P. Labeye, F. Revol, "A micro-opto-mechanical switch integrated on silicon for optical fiber network," *Digest IEEE/LEOS 1996 Summer Topical Meetings*, Keystone, Colorado, Aug. 5-9 (1996), 71-72.

[63] M. J. F. Digonnet, H. J. Shaw, "Analysis of a Tunable Single Mode Optical Fiber Coupler," *IEEE J. of Quantum Electron.* **EQ-18** (4), 746-754 (1982).

[64] G. A. Magel, "Integrated optic devices using micromachined metal membranes," *SPIE Integrated Optics and Microstructures III*, vol. 2686, 54-63 (1996).

[65] Y. W. Kim, M. G. Allen, N. F. Hartman, "Micromechanically based integrated optic modulators and switches," *SPIE Integrated Optics and Microstructures*, vol. 1973, 183-189 (1992).

[66] W. Lukosz, "New Nanomechanical Integrated Optical Devices as Modulators, Switches, and Microphones; Integrated Optical (Bio-)Chemical Sensors," *Micro System Technologies 90*, ed. H. Reichl, VDE-Verlag (1990) 725-732.

[67] P. Pliska, W. Lukosz, "Electrostatically actuated integrated optical nanomechanical devices," *SPIE Integrated Optics and Microstructures (1992)*, vol. 1793, 259-272.

[68] W. Lukosz, "Integrated optical nanomechanical devices as modulators, switches, and tunable frequency filters, and as acoustical sensors," *SPIE Integrated Optics and Microstructures (1992)*, vol. 1793, 214-234.

[69] F. Chollet, M. de Labachelerie, H. Fujita, "Microoptomechanical devices: an electrostatically actuated bending waveguide for optical coupling," *European Symposium on Lasers, Optics, and Vision for Productivity in Manufacturing I*, Besançon, France, June 10-14 (1996).

[70] T. Ueda, F. Kohsaka, D. Yamazaki, T. Iino, "Quartz crystal micromechanical devices, " *Proc. IEEE Micro Electro Mechanical Systems (MEMS 85)*, Philadelphia, PA, Jun. 11-14 (1985), 113-116.

[71] K. Takahata, S. Aoki, T. Sato, "Fine Surface Finishing Method for 3-Dimensional Micro Structures," *Proc. IEEE Micro Electro Mechanical Systems (MEMS 96)*, San Diego, California, Feb. 11-15 (1996), 73-78.

[72] R. A. Miller, G. W. Burr, Y. C. Tai, D. Psaltis, C.-M. Ho, R. R. Katti, "Electromagnetic MEMS Scanning Mirrors for Holographic Data Storage," *Solid-State Sensor and Actuator Workshop*, Hilton Head Island, South Carolina, June 2-6 (1996), 183-186.

[73] J. W. Judy, R. S. Muller, "Batch-Fabricated, Addressable, Magnetically Actuated Microstructures," *Solid-State Sensor and Actuator Workshop*, Hilton Head Island, South Carolina, June 2-6 (1996), 187-190.

[74] M. H. Kiang, O. Solgaard, R. S. Muller, K. Y. Lau, "High-precision silicon micromachined micromirrors for laser beam scanning and positioning," *Solid-State Sensor and Actuator Workshop*, Hilton Head Island, South Carolina, June 3-6 (1996), 33-34.

[75] M.-H. Kiang, O. Solgaard, R. S. Muller, K. Y. Lau, "Surface-micromachined electrostatic-comb driven scanning micromirrors for barcode scanners," *Proc. IEEE Micro Electro Mechanical Systems (MEMS 96)*, San Diego, California, 192-197 (1996).

[76] M. Ikeda, H. Goto, M. Sakata, S. Wakabayashi, "Two Dimensional Silicon Micromachined Optical Scanner Integrated with Photo Detector and Piezoresisitor," *The 8th Int. Conf. on Solid-State Sensors and Actuators (TRANSDUCERS 95)*, Stockholm, Sweden, Jun. 25-29 (1995), **vol. 1**, 293-296.

[77] M. Ikeda, H. Goto, M. Sakata, S. Wakabayashi, K. Imanaka, M. Takeuchi, T. Yada, "Two dimensional silicon micromachined optical scanner integrated with photo detector," *SPIE Micro-Optics/Micromechanics and Laser Scanning and Shaping*, vol. 2383, 118-124 (1995).

[78] Y. Ohtuka, H. Nishikawa, T. Koumura, T. Hattori, "2-Dimensional Optical Scanner Applying a Torsional Resonator with 2 Degrees of Freedom," *Proc. IEEE Micro Electro Mechanical Systems (MEMS 95)*, Amsterdam, the Netherlands, Jan. 29-Feb. 2 (1995), 306-309.

[79] A. A. Yasseen, S. W. Smith, M. Mehregany, F. L. Merat, "Diffraction grating scanners using polysilicon micromotors," *Proc. IEEE Micro Electro Mechanical Systems (MEMS 95)*, Amsterdam, the Netherlands, Jan. 29-Feb. 2 (1995), 175-180.

[80] F. Merat, M. Mehregany, "Integrated micro-opto-mechanical systems," *SPIE Micro-Optics/Micromechanics and Laser Scanning and Shaping*, vol. 2383, 88-98 (1995).

[81] O. Tabata, R. Asahi, N. Fujitsuka, M. Kimura, S. Sugiyama, "Electrostatic driven optical chopper using SOI wafer," *Tech. Digest 7th Int. Conf. on Solid-State Sensors and Actuators (TRANSDUCERS 93)*, Yokohama, Japan, Jun. 7-10 (1993), 124-127.

[82] V. P. Jaecklin, C. Linder, N. F. de Rooij, J.-M. Moret, R. Vuilleumier, "Optical microshutters and torsional micromirrors for light modulator arrays," *Proc. IEEE Micro Electro Mechanical Systems (MEMS 93)*, Fort Lauderdale, Florida, Feb. 7-10 (1993), 124-127.

[83] E. Obermeier, J. Lin, V. Schlichting, "Design and fabrication of an electrostatically drive micro-shutter," *Tech. Digest 7th Int. Conf. on Solid-State Sensors and Actuators (TRANSDUCERS 93)*, Yokohama, Japan, Jun. 7-10 (1993), 132-135.

[84] H. Toshiyoshi, H. Fujita, T. Ueda, "A Piezoelectrically Operated Optical Chopper by Quartz Micromachining," *Journal of Microelectromechanical Systems* **4** (1), 3-9, (1995).

[85] H. Toshiyoshi, D. Kobayashi, H. Fujita, T. Ueda, "A Piezoelectric Quartz Microactuator for a Large Pseudo-Static Displacement," *Jpn. J. Appl. Phys.* **33**, part. 2, no. 12B, L1806-L1808 (1994).

[86] R. Vuilleumier, K. Kraiczek, "Variable entrance slit system for precision spectrophotometers," *Proc. IEEE Micro Electro Mechanical Systems (MEMS 95)*, Amsterdam, the Netherlands, Jan. 29-Feb. 2 (1995), 181-185.

[87] O. Shinchi, H. Hayashida, "Digital Micromirror Device (DMD)$^{\mathrm{TM}}$," *Tech. Digest 9th Meeting of JSPS Committee* No. 136, 2nd Division on Future Machining Technologies, Tokyo, Japan, May 21 (1996), pp.5-10. (in Japanese)

[88] G. Vdovin, S. Middelhoek, "Deformable Mirror Display with Continuous Reflecting Surface Micromachined in Silicon," *Proc. IEEE Micro Electro Mechanical Systems (MEMS 95)*, Amsterdam, the Netherlands, Jan. 29-Feb. 2 (1995), 61-65.

[89] Y. Huang, H. Zhang, E. S. Kim, "Piezoelectrically Actuated Microcantilever for Actuated Mirror Array Application," *Solid-State Sensor and Actuator Workshop*, Hilton Head, South Carolina, June 2-6 (1996), 191-195.

[90] S. Akamine, T. R. Albrecht M. J. Zdeblick, C. F. Quate, "A Planar Process for Microfabrication of Integrated Scanning Tunneling Microscopes," *Sensors and Actuators* **A21-A23**, 964-970 (1990).

[91] S. Akamine, H. Kuwano, K. Fukuzawa, "Development of a microphotocantilever for near-field scanning optical microscopy," *Proc. IEEE Micro Electro Mechanical Systems (MEMS 95)*, Amsterdam, The Netherlands, Jan. 29-Feb. 2 (1995), 145-148.

[92] R. Sawada, H. Tanaka, O. Ohguchi, J. Shimada, S. Hara, "Fabrication of Active Integrated Optical Micro-Encoder," *Proc. 4th IEEE Workshop on Micro Electro Mechanical Systems (MEMS 91)*, Nara, Japan, Jan. 30-Feb. 2 (1991), 233-238.

[93] R. Sawada, "Integrated Optical Encoder," *Proc. 8th Int. Conf. on Solid-State Sensors & Actuators (TRANSDUCERS 95)*, Stockholm, Sweden, Jun. 25-29 (1995), vol. 1, 281-284.

[94] H. Miyajima, E. Yamamoto, M. Ito, S. Hashimoto, I. Komazaki, "Optical micro encoder using surface-emitting laser," *Proc. IEEE Micro Electro Mechanical Systems (MEMS 96)*, San Diego, California, Feb. 11-15 (1996), 412-417.

[95] H. P. Zappe, D. Hofstetter, "A Monolithic Optical Displacement Measurement Microsystem," *Proc. IEEE Micro Electro Mechanical Systems (MEMS 96)*, San Diego, California, Feb. 11-15 (1996), 400-405.

[96] H. P. Zappe, D. Hofstetter, B. Maisenhölder, "Monolithic Optical Interferometers for MEMS Applications," *Digest IEEE/LEOS 1996 Summer Topical Meetings*, Keystone, Colorado, Aug. 5-9 (1996), 35-36.

[97] L.-S. Fan, S. Woodman, "Batch fabrication of mechanical platform for high density data storage," *Proc. 8th Int. Conf. on Solid-State Sensors & Actuators (TRANSDUCERS 95)*, Stockholm, Sweden, June 25-29 (1995), **vol. 1**, 434-437.

[98] W. Tang, V. Temesvary, J. J. Yao, Y.-C. Tai, D. K. Miu, "Silicon Microactuators for Computer Disk Dives," *Jpn. J. Appl. Phys. 35*, part 1, no. 1B, 350-356 (1996).

[99] N. J. Mourlas, K. C. Stark, M. Mehregany, S. M. Philips, "Exploring polysilicon micromotors for data storage micro disks," *Proc. IEEE Micro Electro Mechanical Systems (MEMS 96)*, San Diego, California, Feb. 11-15 (1996), 198-203.

[100] W. Ehrfeld, "The LIGA process for microsystems," *Micro System Technologies 90*, ed. H. Reichl, VDE-Verlag (1990) 521-528.

[101] W. Menz, W. Bacher, M. Hermening, A. Michel,"The LIGA Technique – a Novel Concept for Microstructures and the Combination with Si-Technologies by Injection Molding," *Proc. 4th IEEE Workshop on Micro Electro Mechanical Systems (MEMS 91)*, Nara, Japan, Jan. 30-Feb. 2 (1991), 69-73.

[102] H. Guckel, K. J. Skrobis, T. R. Christenson, J. Klein, S. Han, B. Choi, E. G. Lovell, "Fabrications of Assembled Micromechanical Components via Deep X-ray Lithography," *Proc. 4th IEEE Workshop on Micro Electro Mechanical Systems (MEMS 91)*, Nara, Japan, Jan. 30-Feb. 2 (1991), 74-79.

[103] M. Gerner, T. Paatzsch, L. Weber, H. Schift, I. Smaglinski, H.-D. Bauer, M. Abraham, W. Ehrfeld, "Micro-optical Components for Fiber and Integrated Optics realized by the LIGA Technique," *Proc. IEEE Micro Electro Mechanical Systems (MEMS 95)*, Amsterdam, The Netherlands, Jan.29-Feb.2 (1995), 328-333.

[104] R. Yeh, E. J. J. Kruglick, K. S. J. Pister, "Surface-Micromachined Components for Articulated Microrobots," *IEEE J. Microelectromechanical Systems* **5** (1), 10-17 (1996).

[105] K. Suzuki, I. Shimoyama, H. Miura, Y. Ezura, "Creation of an Insect-Based Microrobot with an External Skelton and Elastic Joint," *Proc. 5th IEEE Workshop on Micro Electro Mechanical Systems*, Travemünde, Germany, February 4-7 (1992), 190-195.

[106] Y. Fukuta, T. Akiyama, H. Fujita, "A Reshaping Technology with Joule Heat for Three Dimensional Silicon Structures," *The 8th Int. Conf. on Solid-State Sensors and Actuators (TRANSDUCERS 95)*, Stockholm, Sweden, Vol. 1, June 25-29 (1995), 174-177.

[107] M. C. Wu, L. Y. Lin, S. S. Lee, K. S. J. Pister, "Micromachined free-space integrated micro-optics," *Sensors and Actuators* **A 50**, 127-134 (1995).

[108] L. Y. Lin, S. S. Lee, K. S. J. Pister, M. C. Wu, "Three dimensional micro-Fresnel optical elements fabricated by micromachining technique," *Electron. Lett.* **30** (5), 448-449 (1994).

[109] S. S. Lee, L. Y. Lin, K. S. J. Pister, M. C. Wu, H. C. Lee, P. Grodzinski, "Passively Aligned Hybrid Integration of 8×1 Micromachined Micro-Fresnel Lens Arrays and 8×1 Vertical-Cavity Surface-Emitting Laser Arrays for Free-Space Optical Interconnect," *IEEE Photon. Tech. Lett.* **7** (9), 1031-1033 (1995).

[110] L. Y. Lin, S. S. Lee, K. S. J. Pister, M. C. Wu, "Micro-Machined Three-Dimensional Micro-Optics for Integrated Free-Space Optical System," *IEEE Photon. Tech. Lett.* **6** (12), 1445-1447 (1994).

[111] S. S. Lee, L. Y. Lin, M. C. Wu, "Surface-micromachined free-space micro-optical systems containing three-dimensional microgratings," *Appl. Phys. Lett.* **67** (15), 2135-2137, (1995).

[112] L. Y. Lin, J. S. Shen, M. C. Wu, A. M. Sergent, "Tunable Three-Dimensional Solid Fabry-Perot Etalons Fabricated by Surface-Micromachining," *IEEE Photon. Tech. Lett.* **8** (1), 101-103 (1996).

[113] D. S. Gunawan, L. Y. Lin, K. S. J. Pister, "Micromachined corner cube reflectors as a communication link," *Sensors and Actuators* **A 46-47**, 580-583 (1995).

[114] M. J. Daneman, N. C. Tien, O. Solgaard, K. Y. Lau, R. S. Muller, "Liner Vibromotor-actuated Micromachined Microreflector for Integrated optical Systems," *Solid-State Sensor and Actuator Workshop*, Hilton Head Island, South Carolina, June 2-6 (1996), 109-112.

[115] J. H. Comtois, V. M. Bright, "Surface micromachined polysilicon thermal actuator arrays and applications," *Solid-State Sensor and Actuator Workshop*, Hilton Head Island, South Carolina, June 3-6 (1996), 174-177.

[116] J. R. Reid, V. M. Bright, J. H. Comtois, "A Surface Micromachined Rotating Micro-Mirror Normal to the Substrate," *Digest IEEE/LEOS 1996 Summer Topical Meetings*, Keystone, Colorado, Aug. 5-9 (1996), 39-40.

[117] T. Akiyama, D. Collard, H. Fujita, "Scratch Drive Actuator with mechanical links for self-assembling of three dimensional MEMS," *IEEE Journal of Microelectromechanical Systems* **6**(1), 10-17 (1997).

[118] M. Oikawa, K. Iga, T. Sanada, "A distributed-index planar micro-lens made of plastics," *Jpn. J. Appl. Phys.* **20**, L51-L54 (1981).

[119] H. Sugiyama, M. Kato, S. Misawa, K. Iga, "Fabrication of Planar Microlens by Transverse Electromigration Method," *Jpn. J. Appl. Phys.* **25** (12), 1959-1960 (1986).

[120] M. Oikawa, K. Iga, T. Sanada, "Distributed-index planar microlens array prepared from deep electromigration," *Electron. Lett.* **17** (13), 452-454 (1981).

[121] J. S. Swenson, Jr. M. Abraham, R. A. Fields, "Fabrication of f/0.7 GaP Microlens Arrays by Accelerated Mass-Transport Smoothing in Sealed Ampoules," *SPIE Micro-Optics/Micromechanics and Laser Scanning and Shaping*, vol. 2383 318-327 (1995).

[122] Z. L. Liau, "Mass-transport fabrication of large-numerical-aperture microopitcs," *SPIE Micro-Optics/Micromechanics and Laser Scanning and Shaping*, vol.2383 310-317 (1995).

[123] H. Sankur, E. Motamedi, R. Hall, W. J. Gunning, M. Khoshnevisan, "Fabrication of refractive microlens array," *SPIE Micro-Optics/Micromechanics and Laser Scanning and Shaping*, vol. 2383, 179-183 (1995).

[124] S. Shirai, T. Serikawa, "Optical Design for Large Scale Free-Space Optical Switching Modules," *IEICE Trans.* **J78-C-I** (5), 231-237 (1995, in Japanese).

[125] E. Pawlowski, W. Fürst, B. Kuhlow, "Micro-optical elements with high efficiency for optical interconnects," *Micro System Technologies 94* (ed. H.Reichl, A.Heuberger), VDE-Verlag Gmbh (1996), 1087-1096.

[126] U. Köhler, A. E. Guber, W. Bier, M. Heckle, "Fabrication of microlenses by plasmaless isotropic etching combined with plastic moulding," *Sensors and Actuators* **A 53**, 361-363 (1996).

[127] W. Karthe, R. Görring, A. Brauer, E. B. Kley, Ch. Wachter, P. Dannberg, T. Possner, "Micro- and integrated optical devices in glasses and polymers for hybrid integration into microsystems," *Microsystem Technologies* **1**, Springer-Verlag, (1995), 59-67.

[128] D. L. MacFarlane, V. Narayan, J. A. Tatum, W. R. Cox, T. Chen, D. J. Hayes, "Microjet Fabrication of Microlens Array," *IEEE Photon. Tech. Lett.* **6** (9), 1112-1114 (1994).

[129] M. Hisanaga, T. Koumura, T. Hattori, "Fabrication of 3-dimensionally shaped Si diaphragm dynamic focusing mirror," *Proc. IEEE Micro Electro Mechanical Systems (MEMS 93)*, Fort Lauderdale, Feb. 7-10(1993), 30-35.

[130] H. Goto, S. Wakabayashi, M. Ikeda, M. Sakata, K. Imanaka, M. Takeuchi, T. Yada, "Microfocusing optical device using piezoelectric thin film actuator," *SPIE Micro-Optics/Micromechanics and Laser Scanning and Shaping*, vol.2383, 136-143 (1995).

[131] O. Solgaard, M. Daneman, N. C. Tien, A. Friedberger, R. S. Muller, K. Y. Lau, "Optoelectronic Packaging Using Silicon Surface-Micromachined Alignment Mirrors," *IEEE Photon. Tech. Lett.* **7** (1), 41-43 (1995).

[132] Y. H. Lo, A. T. T. D. Tran, Z. H. Zhu, G. L. Christenson, "Integrated Micro-Optical Interferometer Arrays," *Digest IEEE/LEOS 1996 Summer Topical Meetings*, Keystone, Colorado, Aug. 5-9 (1996), 25-26.

[133] N. F. Raley, D. R. Ciarlo, J. C. Koo, B. Beiriger, J. Trujillo, C. Yu, G. Loomis, R. Chow, "A Fabry-Perot microinterferometer for visible wavelengths," *Tech. Digest IEEE Solid-State Sensor and Actuator Workshop*, Hilton Head, 170-173 (1992),

[134] T. Geßner, W. Dözel, D. Billep, R. Harn, C. Kaufmann, S. Kurth, "Design, Technology and Characterization of a mirror-array in silicon-micromechanics," *Micro System Technologies 96* (ed. H.Reichl, A.Heuberger), VDE-VERLAG GMBH (1996), 657-662.

[135] J. Bühler, J. Funk, O. Paul, F.-P. Stiener, H. Baltes, "Thermally actuated CMOS micromirrors," *Sensors and Actuators* **A 46-47**, 572-575 (1995).

[136] M. Ross, K. S. J. Pister, "Micro-windmill for optical scanning and flow measurement, " *Sensors and Actuators* **A 46-47**, 576-579 (1995).

[137] D. Collard, T. Akiyama, Y. Fukuta, D. Chauvel, H. Fujita, "Integrated Motion System for Optical Device Alignment and Assembling of Micro-Parts," *Proc. 3rd France-Japan Cong. & 1st Europe-Asia Cong. on Mechatronics*, Besançon, France, Oct.1-3 (1996), 589-593.

[138] H.-J. J. Yeh, J. S. Smith, "Fluidic Self-Assembly for the Integration of GaAs Light-Emitting Diodes on Si Substrate," *IEEE Photon. Tech. Lett.* **6** (6), 706-708 (1994).

[139] S. H. Kravitz, J. C. Word, T. M. Bauer, P. K. Seigal, M. G. Armendariz, "CLASP (Capture and Locking Alignment Spring Positioner) — A Micromachined fiber auto positioning device," *SPIE Smart Structures and Materials 1996: Smart Electronics and MEMS*, vol. 2722, 2-10 (1996).

[140] B. Hillerrich, "Application of Micromachining and Micromechanics to Components for Optical Communication," *Micro System Technologies 90* (ed. H.Reichl, A.Heuberger), VDE-Verlag (1990), 457-464.

[141] G. Nakagawa, K. Miura, K. Tanaka, M. Yano, "Multichannel Fiber Ferrule for a Stable Laser-Diode Array Module," *IEEE Photon. Tech. Lett.* **7** (4), 409-411 (1995).

[142] R. L. Smith, S. D. Collins, "A Wafer-to-wafer Alignment Technique," *Sensors and Actuators* **20**, 315-316 (1989).

[143] B. Acklin, J. Bellermann, M. Schienle, L. Stoll. M. Honsberg, G. Müller, "Self-Aligned Packaging of an Optical Switch Array with Integrated Tapers," *IEEE Photon. Tech. Lett.* **7** (4), 406-408 (1995).

[144] W. Konz, Ch. Vogel, "Micro Optic Components on Silicon Platform with Etched Micromechanical Adjusting Elements," *MICRO SYSTEM Technologies 92*, ed. H. Reichl, VDE-Verlag (1992), 309-312.

[145] M. J.Daneman, N. C. Tien, O. Solgaard, A. P. Pisano, K. Y. Lau, R. S. Muller, "Linear microvibromotor for positioning optical components," *Proc. IEEE Micro Electro Mechanical Systems (MEMS 95)*, Amsterdam, The Netherlands, Jan.29-Feb.2 (1995), 55-60.

[146] Y. Kikuya, M. Hirano, K. Koyabu, F. Ohira, "Micro alignment machine for optical coupling," *Proc. IEEE Micro Electro Mechanicial Systems (MEMS 93)*, Fort Lauderdale, Feb. 7-10 (1993), 36-41.

[147] S. T. Wilkinson, T. W. Kim, N. M. Jokerst, M. G. Allen, "Integration of Thin Film Optoelectronic Devices onto Micromachined Movable Platforms," *IEEE Photon. Tech. Lett.* **6** (9), 1115-1118 (1994).

[148] J. F. Ewen, K. P. Jackson, R. J. S. Bates, E. B. Flint, "GaAs Fiber-Optic Modules for Optical Data Processing Network," *J. Lightwave Tech.* **9** (12), 1755-1763 (1991)

[149] J. W. Parker, "Optical Interconnection for Advanced Processor Systems: A Review of the ESPRIT II OLIVES Program," *J. Lightwave Tech.* **9** (12), 1764-1773 (1991).

[150] D. Uttamchandani, D. Liang, B. Culshaw, "A micromachined silicon accelerometer with fibre optic interrogation," *SPIE Integrated Optics and Microstructures*, vol. 1793, 27-33 (1992).

[151] B. Culshaw, "Fibre optic sensor: integration with micromachined devices," *Sensors and Actuators* **A 46-47**, 463-469 (1995).

[152] J. Marty, A. Malki, C. Renouf, P. Lecoy F. Baillier, "Fibre-optic accelerometer using silicon micromachining techniques," *Sensors and Actuators* **A 46-47**, 470-473 (1995).

[153] A. Méndez, T. F. Morse, "Micromachined Fabry-Perot Interferometer with Corrugated Silicon Diaphragm for Fiber Optic Sensing Applications," *SPIE Integrated Optics and Microstructures*, vol. 1793, 170-182 (1992).

[154] M. A. Chan, S. D. Collins, R. L. Smith, "A micromachined pressure sensor with fiber-optic interferometric readout," *Sensors and Actuators* **A 43**, 196-201 (1994).

[155] M. Ohkawa, M. Izutsu, T. Sueta, "Integrated optic pressure sensor on silicon substrate," *Applied Optics* **28** (23), 5153-5157 (1989).

[156] M. Ohkawa, M. Izutsu, T. Sueta, *Optronics* (1992) no. 9, 97-103 (in Japanese).

[157] M. Ohkawa, M. Izutsu, T. Sueta, "Integrated Optic Accelerometer Employing a Cantilever on a Silicon Substrate," *Jpn. J. Appl. Phys.* **28** (2), 287-288 (1989).

[158] H. Saito, H. Fujita, "Trial Fabrication of a Movable 3-D Optical Waveguide Integrated with Comb-Type Microactuators," *Trans. Inst. Electrical Engineers of Japan* **116-E** (1), 36-37 (1996, in Japanese).

[159] A. A. Boiarski, J. R. Busch, B. S. Bhullar, R. W. Ridgway, V. E. Wood, "Integrated-optic sensor with macro-flow cell," *SPIE Integrated Optics and Microstructures*, vol. 1793, 199-211 (1992).

[160] A. Nathan, Y. Bhatnagar, A. Vadekar, W. Huang, "Fabrication of a silicon Mach-Zehnder interferometer for mechanical measurands," *SPIE Integrated Optics and Microstructures*, vol. 1793, 19-26 (1992).

[161] C. Gorecki, F. Chollet, H. Kawakatsu, F. Fujita, "Silicon-integrated microinterferometer with piezoelectric modulation," *Proc. 3rd France-Japan Cong. & 1st Europe-Asia Cong. on Mechatronics*, Besançon, France, Oct.1-3 (1996), 291-295.

[162] K. E. Burcham, G. N. De Bradander, J. T. Boyd, "Micromachined Silicon Cantilever Beam Accelerometer Incorporating an Integrated Optical Waveguide," *SPIE Integrated Optics and Microstructures*, vol. 1793, 12-18 (1992).

[163] S. Wu, H. J. Frankena, "Integrated optical sensors using micromechanical bridges and cantilevers," *SPIE Integrated Optics and Microstructures*, vol. 1793, 83-89 (1992).

[164] A. M. Young, C. Xu, W. Huang, S. D. Senturia, "Design and analysis of an ARROW-waveguide-based silicon pressure transducer," *SPIE Integrated Optics and Microstructures*, vol. 1793, 42-53 (1992).

[165] T. Storgaard-Larsen, S. Bouwstra, O. Leistiko, "Opto-mechanical accelerometer based on strain sensing by a Bragg grating in a planar waveguide," *The 8th Int. Conf. on Solid-State Sensors and Actuators*, Stockholm, Sweden, vol. 1, June 25-29 (1995), 667-670.

[166] E. Higurashi, H. Ukita, H. Tanaka, O. Ohguchi, "Optically induced rotation of anisotropic micro-objects fabricated by surface micromachining," *Appl. Phys. Lett.* **64** (17), 2209-2210 (1994).

[167] E. Higurashi, O. Ohguchi, H. Ukita, "Optical trapping of low-refractive-index microfabricated objects using radiation pressureflexerted on their inner walls," *Optics Lett.* **20** (19), 1931-1933 (1995).

[168] K. Koyabu, F. Ohira, T. Yamamoto, "Fabrication of Two-dimensional PANDA Fiber Array for Digital Free-space Photonic Switch Module," *Photonics in Switching*, Sendai, Japan, Apr. 21-25 (1996), 136-137.

[169] H. Moilanen, S. Leppävouri, A. Uusimäki, "Laser-induced chemical vapour deposition in piezoresistive pressure sensor fabrication," *Sensors and Actuators* **A 41-42**, 150-155 (1994).

[170] T. M. Bloomstein, D. J. Ehrlich, "Laser deposition and etching of three-dimensional microstructures," *Int. Conf. on Solid-State Sensors and Actuators (TRANSDUCERS 91)*, San Francisco, CA, Jun. 24-28 (1991), 507-511.

[171] M. Alavi, S. Buettgenbach, A. Schumacher, H. J. Wagner, "Fabrication of microchannels by laser machining and anisotropic etching of silicon," *Sensors and Actuators* **A 32** (1-3),299-302 (1992).

[172] M. Mullenborn, H. Dirac, JW. Petersen, S. Bouwstra, "Fast three-dimensional laser micromachining of silicon for microsystems," *Sensors and Actuators* **A 52** (1-3), 121-125 (1995).

[173] K. Minami, Y. Wakabayashi, M. Yoshida, K. Watanabe, M. Esashi, "YAG laser-assisted etching of silicon for fabricating sensors and actuators," *J. Micromech. Microeng.* **3** (2), 81-86, (1993).

[174] K. Ikuta, T. Ogata, M. Tsubio, S. Kojima, "Development of mass productive micro stereo lithography (mass-IH process)," *Proc. IEEE Micro Electro Mechanical Systems (MEMS 96)*, San Diego, CA, Feb. 11-15 (1996), 301-306.

[175] Y. S. Chu, K. Uchino, "Photostrictive effect in PLZT-based ceramics and its applications," *Ferroelectrics* **174** (1-2), 185-196 (1995).

[176] K. Uchino, "Advances in ceramic actuator materials," *Materials Lett.* **22** (1-2), 1-4 (1995).

[177] S. Inaba, H. Kumazaki, K. Hane, "Photothermal vibration of fiber core for vibration-type sensor," *Jpn. J. Appl. Phys.* **34**, part 1, no. 4A, 2018-2021 (1995).

CHAPTER 9

Micromachined Systems for Neurophysiological Applications

Khalil Najafi
University of Michigan
Center for Integrated Sensors and Circuits

CONTENTS

9.1 Introduction / 518

9.2 Motivation and Requirements / 521
  9.2.1 Size / 521
  9.2.2 Power / 522
  9.2.3 Functionality / 523
    9.2.3.1 Multiple transducers / 524
    9.2.3.2 Lead reduction and multiplexing / 524
    9.2.3.3 Amplification and buffering / 525
    9.2.3.4 Other functions / 526
  9.2.4 Stability and reliability / 527
  9.2.5 Biocompatibility / 528
  9.2.6 Cost / 528

9.3 Examples of Implantable Silicon Devices / 530
  9.3.1 Recording microprobes / 530
  9.3.2 Nerve regeneration electrodes / 541
    9.3.2.1 Voltage regulator / 543
    9.3.2.2 Data detector / 544
    9.3.2.3 Preamplifiers / 545
    9.3.2.4 Analog-to-digital converter / 545
  9.3.3 Stimulating microprobes / 548
  9.3.4 Muscular microstimulators / 551

9.4 Packaging Technologies / 555
  9.4.1 Hermetic packaging techniques / 556
  9.4.2 Nonhermetic packaging techniques / 562

9.5 Conclusion and Future Trends / 563

Acknowledgments / 565
References / 565

9.1 INTRODUCTION

Health care and devices used to deliver and monitor health are unquestionably among the most important and critical elements in our society today, and they will continue to become even more important in the future as the world population grows, and as the demand for improved quality and more diverse distribution of health services delivered grows. During the past few decades, we have seen a significant improvement in our understanding of various diseases, in the development of new and exciting tools to fight and detect diseases, and obviously in the cost of delivering health care. Prevention and treatment are equally important as our society moves into the twenty first century, and with these we see an ever-increasing need for better equipment and facilities to assist in delivering the highest quality health care. These equipment need to provide the highest performance at the lowest possible cost. In spite of significant progress having been made in both of these areas, we are now indeed witnessing limitations being imposed on our health care delivery due to very high costs involved. Any new technology that can help reduce this cost will undoubtedly find a significant market and will lay the foundation for the development of a new generation of biomedical systems. One such technology is that of miniaturization and microelectronics-based biomedical devices and systems that are being developed using integrated circuit and micromachining technologies. Solid-state sensors and actuators are a very important component in the development of the next-generation biomedical devices and systems that promise to influence the health care system in ways never imagined before. This chapter focuses on recent advancements made in the development of miniature solid-state biomedical devices and systems for neurophysiological applications for the development of closed-loop neural prostheses.

As solid-state integrated sensors and actuators have reduced in size, have become more reliable, and have come to incorporate on-chip signal processing, their use as research tools as well as clinical instruments in biomedical applications has been increasingly growing. Solid-state sensors and actuators find a wide range of applications in biomedical applications.[1,2] Two general categories can be identified for biomedical devices, invasive and noninvasive (or *in-vivo* and *in-vitro*). Invasive devices are generally either inserted into the biological medium for a short period of time to obtain a measurement, or are implanted for extended periods of time to continuously provide a desirable function. Examples of invasive biomedical devices include catheters used for monitoring and performing surgery on internal organs, or for measuring various biological parameters.[3,4] Noninvasive devices are not inserted into the body and generally interact with the biological environment by monitoring or delivering a number of different media, including X-ray, ultrasound, thermal, optical, etc. Although microelectronics and micromachined sensors and actuators also play a very important role in these applications, this chapter does not address or discuss advances made in this area. It should be mentioned, however, that noninvasive biomedical instruments and devices are obviously very important, and potentially offer the least expensive and traumatic health care delivery option in many applications.

This chapter will concentrate on reviewing the applications, requirements, and examples of invasive solid-state biomedical devices and systems. A majority of these devices are used for sensing and monitoring a number of biological parameters, although solid-state actuators are also finding increasing application for delivering electrical and chemical stimuli to the body.[2] Eventually, it is hoped that a combination of sensors and actuators will be used to realize complete closed-loop biomedical systems that enhance or restore functions for the handicapped, as shown in Figure 9.1. In such systems, the sensors derive control signals from the body and use these signals to generate appropriate stimuli to induce a desired function in the body. The application of the stimuli is adaptively controlled by the feedback signals provided by the sensors. Once developed, these systems will find widespread application in auditory, visual, and muscular prostheses to restore function to the handicapped.[5,6,7]

Applications for implantable biomedical systems are too numerous to discuss in detail here. Generally, two categories for biomedical transducers can be identified: 1) monitoring and measurement of various physical and chemical parameters; and 2) delivery of some stimulus or signal to the body in order to create a controlled change or function.

The first category is rather broad and covers many specific applications. For example, patient monitoring during both surgery and normal daily life is required to ensure that the parameters of interest are in the range they need to be and that the overall health of the patient is stable. These parameters include blood pressure and pulse rate, and chemical species such as glucose and pH level.[8] Another example is the monitoring of electrical biopotentials generated by neurons in order to study and understand the operation of the human nervous system and the organization of complex neural structures. Many devices under this category do not necessarily utilize the measurements obtained to initiate or control another function, as is necessary in prostheses. That is, the eventual objectives for the use of these devices may be just the measurement of one or more biological parameters. Examples of this include measurement of pressure, level of activity, temperature, pulse rate, blood flow, chemical concentrations, etc. Some of these can be performed using external devices, and some require the implantation of a device into the body. For all of these applications, however, one of the most important requirements is that the overall system be as small as possible so that it does not disturb the activity of the patient, unless the individual is restrained to a bed, as in surgery.

In the second category, the main function is the delivery of a stimulus, either chemical, electrical, or mechanical, to a particular part of the body. In order to achieve this, an actuator of some type is required that can deliver the required stimulus. The delivery is typically performed in conjunction with signals provided by a sensor that is either implanted into the body or resides external to the body. An example in this category is the delivery of electrical stimuli to the cochlea in the human ear in order to produce sensation of sound in a completely or partially deaf individual.[6] These stimuli are in the form of current pulses that are delivered through multielectrode probes that are implanted in the scala tympani. Chemical agents such as insulin can also be delivered for those who suffer from diabetes. In this application, the implantable system should be capable of continually monitoring the blood sugar level, compare it to a prespecified level stored electronically in the implant, and deliver a suitable amount of insulin from a local reservoir through microvalves incorporated into the implant. In any of these applications the delivery of the required stimulus is initiated by a control signal that is derived either from the body, such as in closed-loop neural prostheses, or from the external environment, as in an auditory prosthesis that records sound from the external environment and converts it into appropriate electrical pulses for simulating either the scala tympani or the central nervous system.[6]

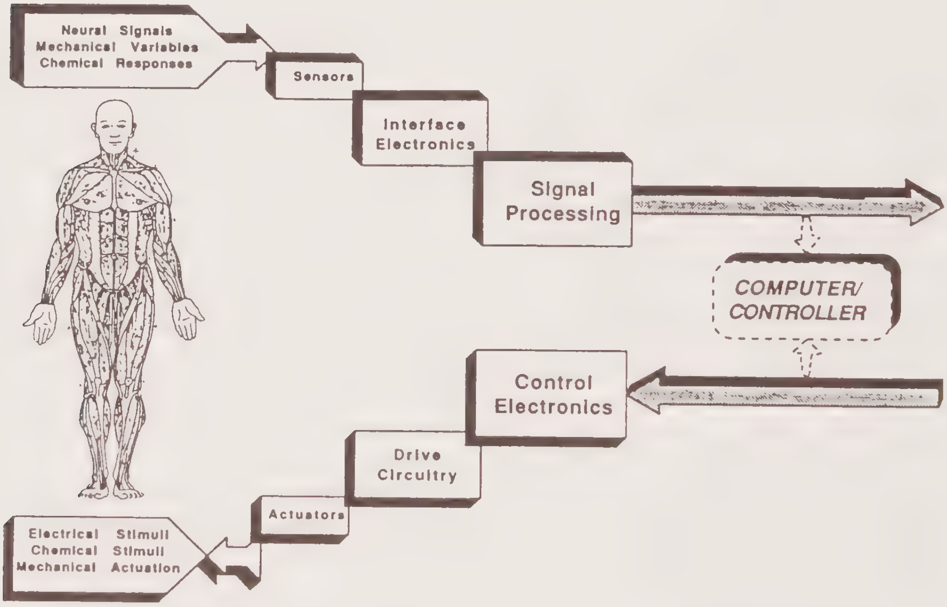

Neural Signals
Mechanical Variables
Chemical Responses
Sensors
Interface Electronics
Signal Processing
COMPUTER/ CONTROLLER
Control Electronics
Drive Circuitry
Actuators
Electrical Stimuli
Chemical Stimuli
Mechanical Actuation

FIG. 9.1 General architecture illustrating typical circuit blocks needed in biomedical systems.

The above two categories are very broad and many other examples can be cited beyond those already presented. Both of the categories, however, require the utilization of systems that can produce the required function, whether sensing or actuation, in a stable and reliable manner. Indeed many of the emerging biomedical implants and devices have to go through long and complex trial studies and clinical tests before they are approved for human use not only to show that they can produce the functions they are designed for, but mainly to ensure that their use does not irreversibly harm the individual. An excellent example of a potentially useful sensor for biological applications is the ion-sensitive field-effect transistor (ISFET),[8,9] used for pH measurement. Although this sensor was developed many years ago, it has been plagued by stability and reliability problems and has not yet been fully incorporated into practical biomedical applications. Therefore, it is worth noting that among the principal requirements for most implantable biomedical systems are reliability and stability. In order to achieve these functions one has to design systems that offer certain features. It is in the implementation of these features where micromachined integrated biomedical sensors and actuators offer their most important advantage.

In this chapter we first present the motivation and requirements for implantable biomedical sensors and actuators, and discuss how and why miniaturization and microelectronics technologies provide a significant improvement over many existing technologies. Next, we present a number of examples of implantable devices, their fabrication, and their potential advantages. Many of these examples are based on research work conducted at the University of Michigan during the past fifteen years. Many other examples of devices developed by other groups exist; however, this chapter focuses mostly on the work performed at Michigan. The chapter then discusses the very important issue of packaging and its significance in implantable biomedical devices. Reliable packages have always been and continue to be a very difficult problem to overcome and a number of new technologies promise to advance our capabilities in this area; these are discussed next. Finally, the chapter presents conclusions and provides a discussion on future trends and challenges.

9.2 MOTIVATION AND REQUIREMENTS

As was discussed above, the need for better and lower cost health care delivery is ever-growing in today's society. Biomedical devices and systems are a very important part of this delivery system, and often limit the achievable performance and drive up the cost. Indeed, biomedical devices and systems face some of the greatest challenges of any device since they have to operate in a very delicate environment, and have to provide the highest reliability possible. In order to understand why solid-state sensors and actuators and their fabrication technologies can provide significant improvement in the performance of biomedical systems, it is important to discuss what requirements are generally demanded of these systems. In the following section, these requirements are briefly described.

9.2.1 Size

The size of any implantable biomedical system has a significant effect on its overall performance, stability, and acceptance by the biological host. In both sensing and actuation applications, the device should be small enough so that it can approach the tissue of interest with minimal damage and couple to the tissue as closely as possible. The lesser the damage to the tissue, the better is the quality of the signals that are recorded from the tissue (in the case of implanted sensors), and the more effective is the impact of the stimuli that are delivered to the tissue. In addition, size plays a very important role in biocompatibility, which plays a central role in the eventual utilization of any implantable device. The body either totally rejects foreign bodies by encapsulating them with fibrous tissue and by generating chemicals that tend to attack it, or reacts violently through infections that eventually damage the area that is under investigation. Reducing the size of an implant can only improve biocompatibility and the overall stability and reliability of the implant.

For some applications, such as in recording of neural signals from the central nervous system, no size is too small, and the system ideally has to have no physical dimensions to it. Obviously this is not possible for implantable systems. However, usually below a minimum size, which differs from application to application, the implanted device is practically transparent to the tissue and can reside within it without any noticeable effect. For example, when recording from the central nervous system, if one can reduce the size of the silicon carrier that supports the recording sites to below about 30 μm, one can effectively make the device transparent to the tissue and minimize disturbance. This is because the dimensions of the biological structures within the cortical tissue (such as the neurons, which are about 30-50 μm in diameter, and blood vessels, which are larger than ~50 μm in diameter) are generally larger than 30μm and any implant smaller than this will cause minimal perturbation to the surrounding tissue. Micromachining technologies offer a significant advantage in this area as they allow many sensors and actuators to be reduced in size without sacrificing the overall performance. In addition, because of the requirement of a very small size, and because of the ever-increasing demand for enhanced system functionality, it is required that in many implantable systems electronic circuitry be integrated with the transducers that interface with the tissue of interest. Although in many research and clinical applications the circuitry can be hybrid attached to the transducer(s), circuit integration inevitably leads to smaller size and improved reliability. In this regard, micromachined sensors and actuators provide the only means to reducing the size while incorporating signal processing electronics adjacent to the transducers. Examples of biomedical devices that offer these features are given later in this chapter.

It should also be noted that one may argue that inclusion of circuitry on an otherwise passive device (with no circuitry) may lead to an increase in size. Although this is correct in general, we should keep in mind that for many present and future biomedical systems there will be some electronic function that needs to be performed as close to the transducer as possible, which in turn demands that this electronic circuitry be integrated with the transducers. In summary, many next-generation biomedical systems will contain a substantial amount of electronic circuitry which should be integrated monolithically with the sensors and actuators in order to reduce the overall size of the system.

9.2.2 Power

Power dissipation and energy consumption are important issues that have to be considered carefully for any device that is to be implanted inside the body. It is desirable to minimize the energy required by a biomedical device for two main reasons. First, it is desirable to minimize the energy requirement of the device, and thus simplifying the task of providing that energy to the device to begin with. For many implantable systems, energy is provided either by a battery or via a wireless link that operates using two antennae (one outside and one inside the body). In both instances the amount of power is limited. In the case of a battery-operated system, the lifetime is also limited, and changing batteries very often is not feasible or practical. Therefore, the device has to conserve energy and has to have as small a power dissipation as possible.

The second issue that needs to be considered as far as power dissipation is concerned is that of safety. There are two aspects to this issue. One is that the voltages and currents that exist within an implantable device should be controlled, and should preferably be small enough not to cause any damage if they are accidentally exposed to the biological tissue. Therefore, it is again desirable to operate devices with low voltages and at low current levels. The second issue is that of heat generation by the device inside the body. An important aspect in the design and packaging of implantable biomedical sensors is that of thermal management and isolation. Any implant should be designed so that its operation will not excessively raise the temperature of the surrounding tissue; otherwise permanent damage will be done. This is particularly important for implants that incorporate internal circuitry. The power dissipated by the circuitry generates heat that should somehow be removed safely by the body. This again requires that low-power circuits be employed wherever possible, which narrows down the choice of different circuit technologies. This is one of the reasons why low-power CMOS is attractive for biomedical applications. In addition to the choice of circuit technology, the fabrication techniques should be capable of providing layers that can couple to the tissue uniformly and appropriately to remove the generated heat in a consistent manner. In this area the designer may not have much choice, but he/she can design the system so that the hotter portions of the device, where the circuitry is located, are remote from the most sensitive regions of the tissue, where the transducers are located.

In both of these instances, it is clear that micromachined solid-state sensors and actuators have a lot to offer. Their small size and many of the transduction techniques used in them are inherently capable of supporting very low power dissipation. In addition, because electronics can be incorporated into the package of many of these devices, and since this circuitry can be designed for low power operation, the issue of thermal management and heat generation are either nonexistent or can be easily handled by proper packaging. Examples provided later in this chapter illustrate these features of solid-state sensors and actuators.

9.2.3 Functionality

The goal of any biomedical system is to perform one or more functions when it is interfaced with the biological medium. The number of these functions and the manner in which these functions are performed eventually determine the usefulness of the system. It is in the delivery of higher performance and added functions that micromachined sensors and actuators offer their most significant advantage over existing systems. Emerging applications for biomedical systems require that many measurements and functions be achieved with a single device, whether it is a catheter that is inserted into the body or an implant that is to produce a particular action in the biological environment. Because of their small size and their versatility, solid-state transducers are capable of delivering such features. It is possible to measure a number of biological parameters, such as blood pressure and flow, temperature, and chemical species all in a single small device. In order to accomplish this, solid-state transducers offer two capabilities, the availability of more than one sensor type on a single carrier, and the possibility to incorporate electronics to improve the performance.

9.2.3.1 Multiple transducers

Measurement of multiple biological parameters in-situ is very important to many biomedical applications. For example, in patient monitoring it is required that all vital health signs, such as pulse rate, pressure, temperature, and chemical composition of the biological medium be measured accurately and as minimally invasive as possible. An application requiring this is in heart surgery where multipoint measurement of blood pressure inside an artery together with body temperature and heart rate are desirable.[4] While some of these measurements can be performed externally, some others have to be made at the point of use. In addition to measuring a number of different parameters, it is also desirable that the same parameter be measured or a stimulus be delivered at multiple locations inside the body so that either a distribution map can be constructed or redundant measurements are obtained to improve reliability. In both instances, solid-state sensors and actuators are invaluable and indispensable.

In addition to sensing and actuation functions performed by the transducers, often the signals produced by these devices have to be processed. The signals recorded by many biomedical sensors have to be sent out to the outside world for further processing and storage. These signals are typically very small and the sensor interface typically has a high impedance. In addition, for many implantable transducers it is desirable to record or deliver many signals using a single packaged device. Therefore, the number of leads that need to be attached to the device can become large. Inclusion of interface electronics and signal processing circuitry at the transducer site can help to alleviate these problems. In the following sections, some of the main functions and features of signal processing and interface circuitry in biomedical systems are discussed.

9.2.3.2 Lead reduction and multiplexing

Reducing the number of output leads is an important function for a majority of implantable biomedical systems. External leads are not only a major source of failure but are also relatively large compared with the device itself. Data multiplexing can not only reduce the amount of circuitry required for the sensor, but also reduces the number of external leads by either multiplexing several sensed signals onto a single output lead, or by superimposing clock and control signals over the power lines.[4,10] Low-noise analog multiplexing is especially an important function in sensing arrays and systems which require the simultaneous measurement of many signals in order to accurately measure the parameter of interest. In addition, reducing the number of leads that are interconnected to a sensor package is very important in simplifying the packaging, reducing the cost, and improving the long-term reliability of the system in which the sensor is integrated. In addition to sensing applications, data multiplexing can significantly reduce the number of leads required in an actuating system. In many such applications, often several actuators and delivery channels are incorporated into a single implantable package. Multiplexing allows several control signals, along with clock and power, to be delivered sequentially into the implant and latched for later delivery into biological tissue according to a prespecified protocol.

It should be noted that while multiplexing goes a long way in reducing the number of leads, it does not reduce it to zero. It is of course desirable, and often needed, that there be no leads at all. This can be done through telemetry of both signals and power to the implant.[11] Indeed, implantable telemetry systems are becoming more important, and feasible, as new technologies are developed to practically incorporate them into the overall system. The incorporation of telemetry systems into implantable biomedical systems can significantly enhance the overall reliability of the system and will simplify the packaging. It is anticipated that before the end of the century totally implantable telemetry-operated prosthetic systems will be developed for both patient monitoring and prostheses. In fact, most auditory implants in patients today utilize some form of telemetry for either data or power transfer across the skin. One important requirement for the development of telemetry system is low power dissipation. The microelectronics industry is intensively pursuing the development of a variety of low-power techniques for next generation systems. The results of this development effort will indeed be directly applicable to the design of wireless implantable systems. Another requirement is reduced susceptibility to interference caused by the rather noisy environments existing in hospitals and homes. This problem, however, is not as severe in many biomedical applications because the wireless link between the implant and the external transmitter is very short and the internal and external systems are closely coupled. In addition, the effects of external interference can be minimized using proper shielding and packaging and by employing appropriate signal processing and encoding techniques which enhance the reliability of the overall system.

9.2.3.3 Amplification and buffering

The signals recorded by many sensors are typically low in amplitude, and the sensor interface usually exhibits a high impedance at the frequencies of interest. Integration of interface electronics and signal processing circuitry at the sensor site serves a number of functions, including signal amplification, impedance transformation, and signal filtering and buffering.

Amplification of recorded biological signals remains one of the most important functions in many applications where the amplitude of the signal is typically low. Amplification of these signals at the sensor site before transmission to the outside world not only enhances the overall signal-to-noise ratio, thus reducing the effects of noise from the environment, but also allows for the full utilization of the dynamic range of an analog-digital converter for those sensors that incorporate an analog-digital converter (ADC) in the sensor module. For many integrated sensors this amplification can be achieved by using MOS and bipolar amplifiers that require only minimal amount of circuitry with nominal gain, bandwidth, and performance specifications. CMOS amplifiers are perhaps the most suitable since they provide high gain and high input impedance through a relatively simple and compact circuit, and are readily compatible with integration of high-density digital circuitry on the same chip.[12] CMOS amplifiers are now easily capable of providing open-loop gains of 90 dB, are fast, can be compensated for their offset using various techniques involving charge storage and cancellation, and consume little power and die area. These amplifiers are 3-5 times smaller than bipolar amplifiers, making it possible to include tens of per-channel amplifiers on a single chip. In addition, the input impedance of CMOS amplifiers is very high, which makes them ideal for sensor applications. Although MOS amplifiers have typically had larger noise levels than their bipolar counterparts, various circuit and fabrication techniques are used to minimize the noise in applications where the noise level of the interface electronics determines the resolution of the sensing system.[12] More sophisticated amplifiers incorporate programmable gain capability to maximize signal-to-noise ratio and optimize the amplifier output range for analog-digital converters.

Since many implantable sensors have characteristic impedances that are very high (in the megaohm range), it is required that the output leads are buffered to protect against leakage and parasitics in the face of long-term exposure to body fluids. Providing this enhanced drive capability for the output leads is crucial to the long-term stability of the overall implanted system and simplifies the packaging and encapsulation constraints on the output leads. Often the amplification and impedance transformation can be achieved simultaneously using the same amplifier. For some applications where gain is not immediately necessary on the sensor chip, a simple buffer circuit can be integrated in the form of source followers that require only a few transistors and consume very little area and power. For capacitive-based sensors, one has to ensure that the input capacitance of the buffer is as small as possible to minimize capacitive loading. Signal filtering can also be incorporated in amplifiers. For multiplexed multisensor systems, filtering of the signal is often required to prevent such problems as aliasing that can introduce high frequency noise into the signal passband, thus overriding the low-amplitude sensed signal. Signal filtering can also improve the signal-noise ratio of the sensor as it filters the out-of-band high frequency noise that may be introduced by the electronic devices themselves. A variety of low-pass, bandpass, and high-pass filter circuits are available using switched-capacitor circuit techniques or more conventional resistor-capacitor techniques.

9.2.3.4 Other functions

In addition to these primary functions, electronic circuitry is increasingly used to perform functions such as self-testing of the transducers and the circuit blocks, analog-digital conversion, offset cancellation to allow full utilization of the dynamic range of the ADC, radio frequency telemetry for wireless communication and power transfer, and on-chip data processing and interpretation. Reducing the data processing overhead on the external computers by signal processing at the sensor site will be required for many future implantable sensors. Biological events have a very low bandwidth and it is often not necessary to transmit the full passband to the outside world. For many applications having knowledge of the occurrence of an event or of the frequency of the occurrence of an event is all that is needed. In these application it is very desirable to perform event detection and discrimination at the sensor site. For example, in closed-loop neural prostheses where the control signals needed for operating the prosthetic device are derived from the neural signals recorded from the nervous system, the useful information is often included in the rate of neural discharges rather than the absolute amplitude of each discharge.[13] Therefore, one can use an on-chip processor that transmits only this information and discards the rest. This can effectively increase the bandwidth of the transmission link and can reduce the overhead of the external signal processing units. Since most of the data in this form can be transmitted digitally, the external transmission protocol can be designed around a bus-organized system, thus greatly increasing the data transfer speed and improving the overall signal-to-noise ratio. It is believed that for many future closed-loop prosthetic systems such signal processing has to be performed at each sensor site.

It should be again emphasized that signal amplification, filtering, and multiplexing of several data channels onto a single data line also reduce packaging problems and inherently improve the reliability of the device by enhancing the signal-noise ratio and by eliminating the need for too many external leads.

9.2.4 Stability and Reliability

Of all the requirements imposed on biomedical devices, reliability, safety, and stability are the most important. Any biomedical device has to be tested and proven to be safe and it should provide stable and reliable performance over the life of the device. Because of these very hard requirements, biomedical systems are generally much more expensive than other similar systems used in different applications. It is these two requirements of reliability and stability, in addition to others such as sensitivity and resolution, that make micromachined transducers particularly suitable and indeed required in biomedical applications. Achieving these two specific requirements is made somewhat easier by using micromachined transducers through several means. First, micromachined transducers are generally more stable than their discrete counterparts. This is due to the improved mechanical and electrical stability of the materials (such as silicon and other deposited thin films used in integrated circuits) used in their construction. Second, reliability and stability can both be significantly improved by incorporating redundancy into the system. Micromachined technologies allow the formation of large arrays of sensors and actuators in a small area. The outputs of these sensors, some of which measure the same parameter as other sensors in the device, can be averaged, processed, and used to make sure that they all have measured the parameter of interest within a specified range compared to other similar sensors in the array. In addition, by using redundancy one can ensure that if one sensor fails there exist similar sensors to provide the required measurement. Third, stability and reliability can be improved to some degree by integration of signal processing. The features provided by signal processing circuitry have been discussed above. In addition to the above, micromachined devices provide improved stability and reliability because of their smaller size (less disturbance to the tissue), lower power dissipation, and in some instances better packaging and encapsulation techniques. Considering all of these potential advantages, and considering the fact that many biomedical systems are becoming ever more complex, it is obvious that micromachined solid-state sensors and actuators will continue to play a bigger role in the development of next-generation biomedical systems. Among the greatest features they offer are improved reliability, stability, and safety.

9.2.5 Biocompatibility

Any technology and material used for the fabrication of an implantable biomedical device and its package should be biocompatible with biological tissue and should be able to tolerate exposure to the harsh environment inside the body. The issue of biocompatibility cannot be emphasized enough since it is the only requirement that is unique to this application. Because of biocompatibility issues, it is often not possible to use established materials and technologies. For example, many hermetic packaging technologies that are routinely used in the IC industry cannot be used in biomedical devices because of the reaction of biological tissue to them.[14] This incompatibility exacerbates the problem as it limits the range and type of materials that can be used in biomedical systems. Fortunately, many of the standard materials used in micromachined transducers, such as silicon, silicon dioxide, silicon nitride, etc., have been found to be biocompatible and easily accepted by the tissue.[15,16] In addition to material biocompatibility issues, the size of a device also affects its biocompatibility and tissue acceptance. As mentioned before, solid-state transducers offer a distinct advantage in this area over discrete systems because of their significantly smaller dimensions. In either case, it is believed that micromachined transducers and biomedical systems generally offer a significant improvement in terms of biocompatibility and long-term stability inside a biological medium. New technologies and materials are also being constantly developed to promote biocompatibility by coating surfaces of implantable systems with different biocompatible materials. It is certain that we will continue to see a major effort being exerted on addressing the issue of biocompatibility for any device that is introduced into the body. Micromachined biomedical devices will benefit from advances made in this area.

9.2.6 Cost

As discussed before, any new technology that can reduce the cost of biomedical systems and the delivery of health care will make a significant impact on our society. By reducing the cost and size of biomedical systems, several goals can be accomplished. First, and foremost, health care can be provided to a larger segment of the society to help both treatment and prevention. Second, many biomedical systems can be made disposable, which can significantly reduce the cost due to sterilization and the danger of cross contamination. Third, by reducing their size and cost, more biomedical devices can become portable and treatment can be delivered outside of the hospital or physician office, thus reducing the cost. It is clear that meeting all of these objectives can be more easily accomplished using micromachined solid-state transducers. These devices are small, and because of the batch fabrication technologies used in their manufacturing, they can be very low cost at high volumes. Disposability is an important feature for many biomedical applications because it truly diversifies the delivery of health care and takes it beyond the walls of hospitals, which are traditionally one of the most expensive components of our health care system. An excellent example of the enormous potential of micromachined systems under intense development is the concept of single-chip chemical and biological analysis systems that can be used to analyze blood, measure different biological species, and provide physicians and patients with instant results.[17] These micromachined "lab-on-chip" systems will revolutionize the field of biomedical systems since they have the potential of reducing the cost, the size, and the complexity of biological analysis by several orders of magnitude and in so doing distribute health care delivery to an extent never seen before. It is beyond the scope of this chapter to discuss these systems. In the next few years we will witness the development of miniaturized biomedical systems that we could not even iamgine until a few years ago.

Although low cost is a critical factor in the success of a biomedical device, it is not always the determining factor in choosing a suitable technology. Rather, overall performance, and particularly reliability and stability, are more important for biomedical applications. These performance parameters are often affected by the packaging technique used, and anything that can be done to simplify packaging will often lead to lower overall cost. Implanted systems are required to operate in the body for long periods of time, some for many decades, and it is extremely difficult, costly, and often dangerous, to perform surgery to replace the failed system. Therefore, the determination of cost in these systems should always take these issues into account. While total device cost is still important in biomedical applications, it is not the driving factor.

In summary, implantable biomedical devices and systems have to satisfy a number of requirements before they can be successfully used in real-world applications. Many of these requirements can be satisfied using microfabrication and micromachining technologies. Solid-state sensors and actuators have been under development for some twenty years and some devices have already had major impact on the research and understanding of biological tissue as well as commercial applications. The use of these devices and micromachined technologies will continue to grow in the future. In the following section a number of examples of devices that researchers at the University of Michigan have developed during the past fifteen years are presented. The majority of these devices have been developed for neurophysiological applications and for closed-loop neural prostheses for restoring function to those who have lost function in some part of their body. Prosthetic systems usually require sensors to record signals from the body, signal processing electronics to process these complex biological signals and generate control commands, and actuators with which control signals can be delivered to the body. Therefore, both sensors and actuators are typically needed in these systems.

9.3 EXAMPLES OF IMPLANTABLE SILICON DEVICES

Although there exist a wide variety and range of applications for biomedical devices, this chapter presents a brief review of a number of devices that have been developed for use in implantable biomedical devices for use in neural-based prostheses. The primary goal in any prosthetic device is to restore function to a disabled part of the body using devices that are designed to bypass the dysfunctional organ. Some prostheses incorporate artificial organs that are attached to the body, such as artificial arms and legs. Most of these devices are passive and are operated by the individual and in many instances are simply attachments to the body. However, it is widely believed that it is possible to utilize the existing musculature of the body and control these structures using electrical stimuli to activate them and perform useful and controllable tasks. In order to accomplish this, the sense and control channels that naturally exist within the intact human system need to be restored and communication between the central nervous system and the organs that have lost their connection to the brain need to be reestablished. This is illustrated in Figure 9.2, which shows the primary blocks used in such a closed-loop neural prosthetic device. In these active prostheses, electrical signals generated by brain cells are recorded using miniature probes, the signals are processed to generate control signals, and these signals are in turn used to drive stimulators that directly interface with the disabled organ to restore some function to it. Although Figure 9.2 shows the example for the case of neural recording and muscular stimulation, it should be mentioned that the stimulated organ need not be necessarily a motor organ. In fact, it can be another neural structure, such as those existing in the ear or the eye in the case of auditory or visual prostheses. In either case, the loop is completely closed and operation of the system is ideally transparent to the user. That is, the recipient of the system would not know the difference between this artificial system and the natural system. In order to develop such systems, implantable devices that can perform the functions of recording (sensing), signal processing, and stimulation (actuation) need to be developed. Furthermore, the overall system has to be biocompatible, stable, and should last in the body for an extended period of time. The following examples present devices that have been developed at Michigan to provide the required functions.

9.3.1 Recording Microprobes

The recording of biopotentials generated electrochemically by individual nerve cells in the central and peripheral nervous systems has been one of the main techniques used to enhance our understanding of how the brain processes information to control all basic bodily functions, of the basic structures of neurons and neural systems, and of the interaction between neurons. In addition, the recording and processing of control signals from many neurons is also required for the successful development of closed-loop neural prosthesis.[5] Most prosthetic devices depend on neuro-electronic recording and stimulation to drive nerve or muscle fibers that are no longer supplied with inputs from their normal sensors or control channels. The recording of the required neural signals from both individual neurons and a population of neurons is typically performed using recording microelectrodes that are intrusively implanted into the biological tissue of interest. The quality of the recorded neural signals depends heavily on the structure and characteristics of the recording probes used. The enormous complexity of the central nervous system, together with a lack of proper investigative tools, has prevented neurophysiologists from finding the answers to many questions, and from verifying many theories related to the brain. Therefore, a great deal of effort has been expended during the past few decades on the development of suitable recording microprobes and instrumentation tools to allow long-term, stable, and high-quality recording of neural biopotentials from both the central and peripheral nervous systems.

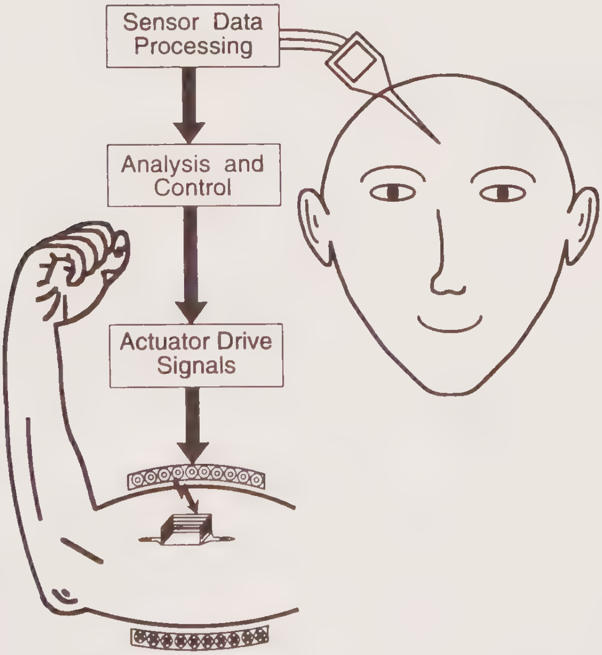


FIG. 9.2 Illustration of a closed-loop neural prosthesis. Signals are recorded from the brain, and then processed and used to stimulate another part of the body.

One of the most widely used techniques for studying the central nervous system at the cellular level is the recording of extracellular biopotentials generated electrochemically within individual neurons. When a neuron receives sufficient stimuli from other cells, its cell membrane depolarizes, causing ionic currents to flow in the vicinity of the cell. The voltage drops associated with this extracellular current can be sensed if a suitable probe is inserted into the vicinity of the neuron. Typically an extracellular single-unit action potential is a spike of about ±50-500 μV in amplitude and has a frequency content from 100 Hz to ~6k Hz. In order to record these action potentials, the probe must penetrate the narrow clefts between cells and approach the active neuron without irreversibly damaging the neuron or the cells driving it. Hence, a most important issue in the design of any recording microprobe is its size, which should be at worst comparable to the neurons, which are usually 50 μm or less in diameter. In practice, however, the probe should be as small as possible and have a shape that facilitates entry and movement through tissue. The small size of the microprobe is also important since it allows the probe to approach and couple to the target cells as closely as possible, thus improving the signal-to-noise ratio of any recorded activity. In addition to the physical requirements, there are several electrical characteristics that a recording microprobe should satisfy. The recording sites should be stable against long-term exposure to biological fluids, and should be capable of recording the electrical activity of the neurons with minimum noise. Because of their low signal levels, the recorded neural signals are often amplified using off-chip, or on-chip amplifiers. One of the most promising approaches used in the development of recording microprobes and biomedical sensors involves the use of integrated circuit microfabrication and silicon micromachining.

Integrated circuit and silicon micromachining technologies have been used to produce multichannel recording and stimulating microprobes for use in a variety of electrophysiological applications.[10,18] The basic structure of a silicon microprobe is shown in Figure 9.3.[18] The probe consists of a precisely micromachined silicon substrate that supports an array of thin-film conductors. These conductors interconnect the recording/stimulating sites along the shank of the probe with signal processing circuitry located at the rear of the substrate. These conductors are insulated above and below by insulators, with openings defined photolithographically at the sites and the output pads.

The fabrication sequence for a multielectrode passive probe structure is shown in Figure 9.4.[18] The substrate material is a silicon wafer of <100> orientation, and desired impurity concentration and thickness determined by the requirements for the integration of on-chip circuitry for signal processing. The wafer is first thermally oxidized and the oxide is removed selectively using photolithography to expose silicon areas in the shape of the intended probe. This shape can be arbitrary and is not dependent on wafer crystallography. The wafer is next subjected to a deep boron diffusion using the thick oxide area as a selective mask. The exact time and temperature of diffusion are selected to set the thickness of the final probe substrate. This diffusion results in a boron concentration approaching 10^{20} cm^{-3} at a depth of about 15μm into the silicon.

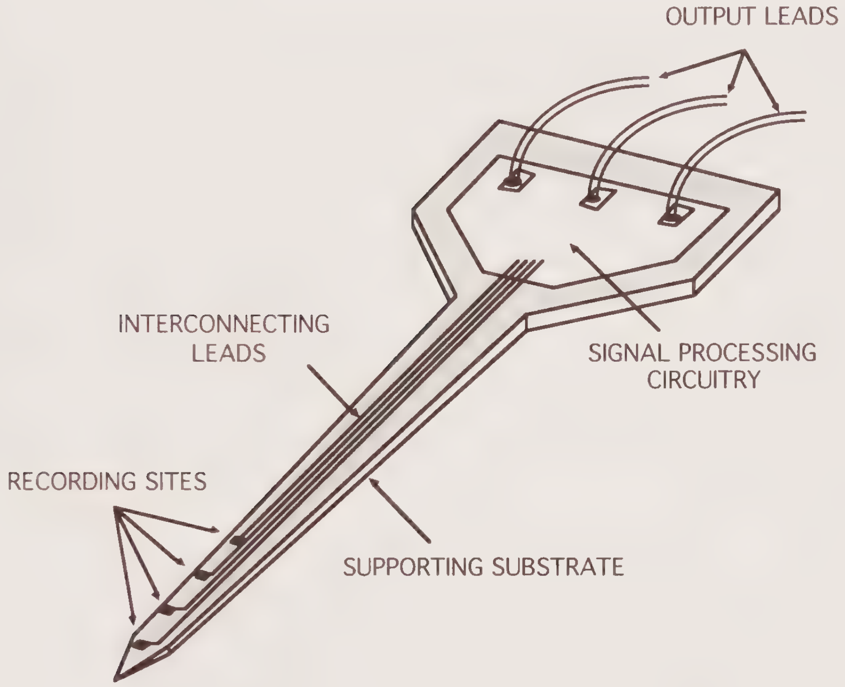

OUTPUT LEADS
INTERCONNECTING LEADS
SIGNAL PROCESSING CIRCUITRY
RECORDING SITES
SUPPORTING SUBSTRATE

FIG. 9.3 Structure of a multichannel silicon recording microprobe.[18]

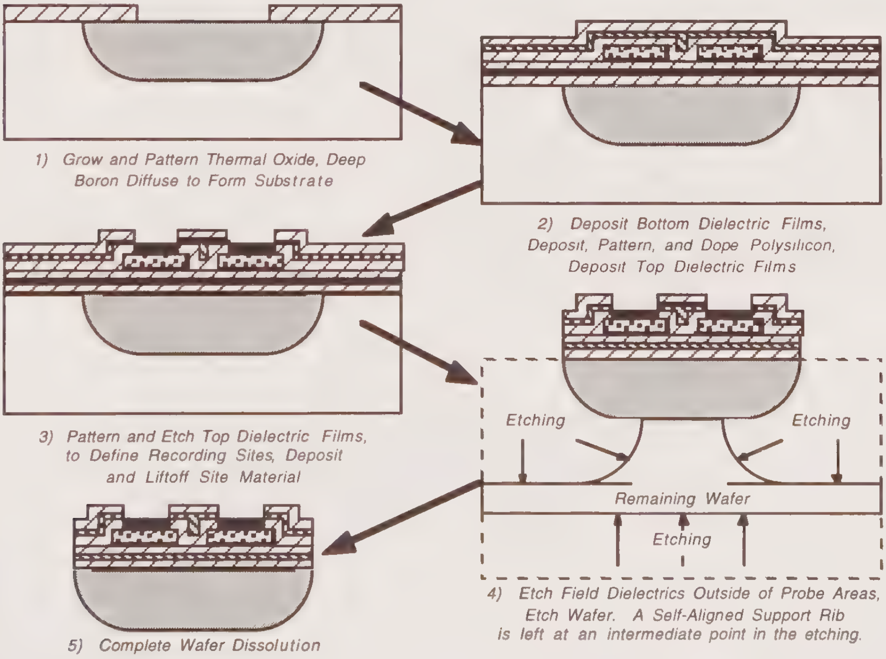

1) Grow and Pattern Thermal Oxide, Deep Boron Diffuse to Form Substrate
2) Deposit Bottom Dielectric Films, Deposit, Pattern, and Dope Polysilicon, Deposit Top Dielectric Films
3) Pattern and Etch Top Dielectric Films, to Define Recording Sites, Deposit and Liftoff Site Material
Etching
Etching
Remaining Wafer
Etching
4) Etch Field Dielectrics Outside of Probe Areas, Etch Wafer. A Self-Aligned Support Rib is left at an intermediate point in the etching.
5) Complete Wafer Dissolution

FIG. 9.4 Fabrication sequence for passive multielectrode microprobes.[18]

Following boron diffusion, the wafer is again thermally oxidized and additional dielectrics are also deposited on the wafer using chemical vapor deposition (CVD). This combined oxide-nitride-oxide dielectric serves as the insulating layer between the supporting substrate and the thin-film electrode conductors, and it therefore controls capacitive coupling to the substrate. The low resistance of the boron-doped silicon substrate ensures that the final probe exhibits negligible crosstalk between adjacent electrodes since the substrate serves as a ground plane under the electrodes.[19] The triple dielectric structure results in a composite insulator that is in a neutral mechanical stress condition so that the resulting probes will be flat and not warped due to the differing thermal expansion coefficients of the materials used. The material used for forming the thin-film interconnects between the recording sites and the output bonding pads is deposited at this time. Typically, polysilicon is used since it is readily deposited by CVD and can be oxidized directly to silicon dioxide to provide a smooth transition from conductor to insulator. In order to reduce its sheet resistance, it should also be doped, which can be achieved with boron or phosphorus. The polysilicon is then patterned after deposition using photolithography and reactive ion etching. The upper dielectrics, which are another combination of silicon dioxide and silicon nitride, are then deposited using CVD. At this point, another mask is used to selectively define photoresist openings over the insulated recording sites and the output bonding areas. Openings are etched through the upper dielectrics to expose the interconnect material. With the resist still in place, these surfaces are etched free of any residual oxide and are coated with successive layers of chromium-gold or titanium-iridium; notice that any metal can be deposited for other applications that require different materials. The resist is then removed through a lift-off process, thereby removing the overlaying metal from the wafer everywhere except in the recording site and bonding pad areas. Thus, the recording sites are inlaid with gold in a single process that is self-aligned, i.e.; the gold inlays and the dielectric openings are produced in the same masking step, avoiding the need for any special registration. Next a final photoengraving step is used to remove both the upper and the lower dielectric layers from areas of the chip outside the area of the intended probe.

This completes the processing of the probe with the exception of its separation from the wafer. This is typically done in two steps. First, the wafer is thinned using a back-side unmasked etch in a hydrofluoric-nitric acid mixture.[18] Next, the wafer is placed in an anisotropic silicon etch composed of ethylenediamine, pyrocatechol, and water (EDP),[20] at a temperature of 115°C. This etchant attacks undoped silicon but does not attack any of the other materials used in the probe structure. In particular, it also does not attack boron-doped silicon where the boron concentration exceeds 5×10^{19} cm^{-3}.[21,22] Thus, the anisotropic etch proceeds from both sides of the wafer until the wafer has been completely removed, leaving only the boron-doped silicon substrates and their dielectric/conductor patterns. Finally, the released probe structures are recovered from the solution and rinsed.

The probe process described above requires only four masks and is capable of high yields. Based on optical inspection, yields have typically exceeded 80 percent. The process is compatible with wafers of standard thickness and does not require double-sided processing. All etching steps are highly selective and self-stopping. Figure 9.5 shows an SEM photograph of the shank of a silicon microprobe fabricated using the above process. The probe is well-defined and can have any desired two-dimensional shape. The substrate thicknesses range from 8 to 15μm, depending on the boron diffusion schedule used. Typical shank lengths have been 1.5–3mm with shank widths normally tapered from a width of 10μm or less near the tip to 100μm near the back. All the features of the probe can be controlled to within 1μm or better. These probes are strong yet flexible and can penetrate intact pia-arachnoid layer with no noticeable bending and minimal dimpling of the cortical surface.[23] The substrates are capable of being elastically bent through angles as great as 90° without breaking.

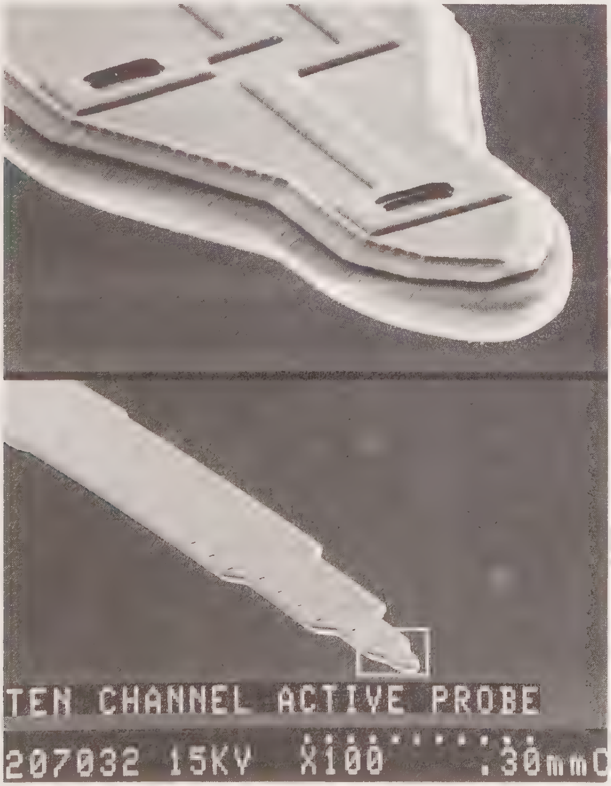


FIG. 9.5 SEM photograph of a multichannel silicon microprobe.

In addition to recording neural activity in two dimensions (i.e., in-depth and laterally), as is possible by the electrodes shown above, an important goal for many next-generation central nervous recording systems is the development of a chronic structure capable of recording three-dimensional activity throughout a volume of tissue. Figure 9.6 shows a three-dimensional multielectrode system developed at the University of Michigan.[24,25] Similar three-dimensional systems have also been developed at the University of Utah.[26] In the Michigan's three-dimensional system, several 32-electrode active multishank probes (to be discussed later) are precisely positioned in a micromachined silicon platform that floats on the surface of the cortex. The platform forms an ideal base for the use of hybrid circuitry for additional signal processing and, possibly, for telemetry. The micromachined platform is formed using deep boron diffusion, similar to the probes themselves. The probes are inserted through slots that are formed in the platform; the slot width and probe thickness are designed to be within 1–2 μm. Spacer bars are also micromachined from silicon and are placed orthogonally in the slots in the probes to keep them parallel in relation to one another during assembly. Electrical interconnects are transferred between the individual probes and the conductors on the platform using nickel electroplating technology as described in reference 24. This process allows for multiple low-resistance lead transfers to be achieved. Each probe shank in the above system supports four recording electrodes, to result in a total of 64 recording electrodes system. If active probes were used, each shank could easily support eight recording sites to result in a system with up to 256 recording sites.[25]

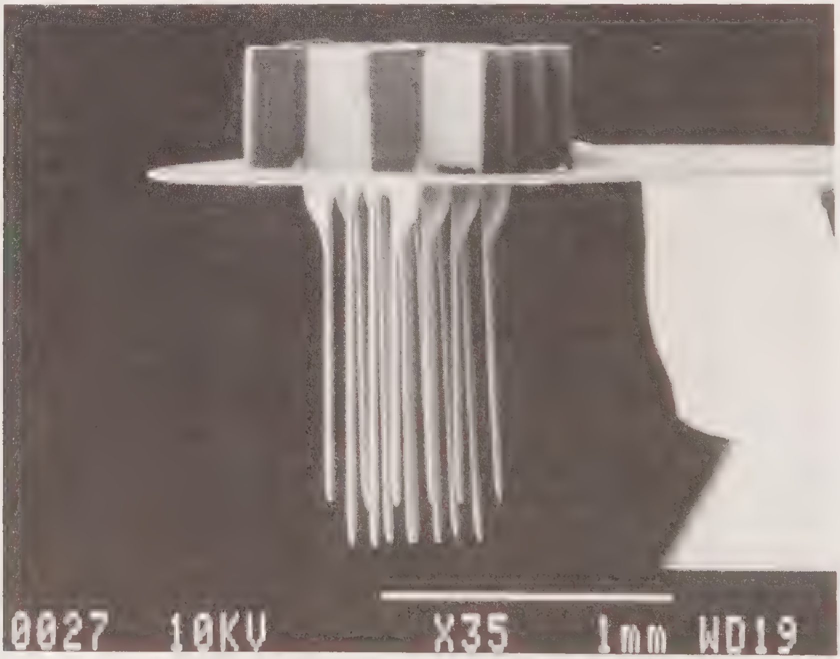


FIG. 9.6 A three-dimensional multielectrode system developed at the University of Michigan.[24,25]

For the successful operation of this multielectrode probe over the long term, signal processing electronics should be directly integrated on the probe substrate. This minimizes the number of output leads, boosts the signal amplitudes, and reduces the sensitivity to the effects of leakage from the output leads, which are the most leakage-prone portion of the system. In addition, electronics can be used to change the output data format, provide on-chip stimulation capability, and allow for remote testing. The inclusion of on-chip circuitry becomes even more important for prosthetic applications where the probes will have to be interfaced to microcomputers for processing of control signals.

Figure 9.7 shows the overall system organization of the electronics for an active multichannel recording electrode system.[10] The system consists of two circuit portions: 1) implantable electronics for integration directly on the probe substrate; and 2) external signal processing electronics for demultiplexing and reconstructing the multiplexed neural signals. The on-chip electronics consists of per-channel preamplifiers, followed by an analog multiplexer and a broadband output buffer to drive the external data line; the circuitry is powered by a regulated battery. The implantable circuitry requires only three output leads for power, ground, and data. In order to allow external regeneration of the on-chip sample clock, a synchronization pulse is inserted once each frame as an eleventh channel. The external electronics strips off these pulses, generates the sample clock, and demultiplexes neural signals for external recording/processing equipment.

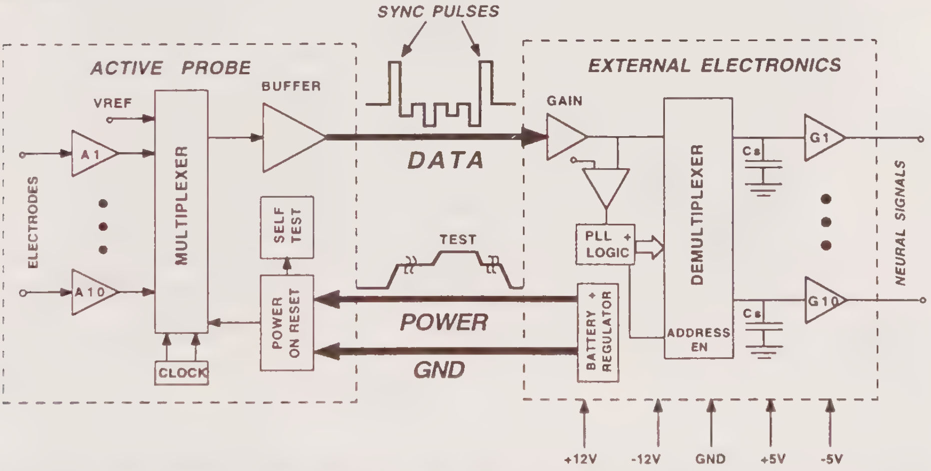


FIG. 9.7 Overall system organization of the electronic system for active multielectrode microprobes.[10]

Figure 9.8 shows a detailed block diagram of the on-chip implantable electronics for use in a ten-electrode active probe.[10] The recorded neural signals from ten electrodes are first received by on-chip preamplifiers that boost the signal levels. These preamplifiers are internally compensated to provide a bandwidth of 15 kHz in order to reduce the total system noise below 20 μVrms. The amplified neural signals are then multiplexed onto a single data line using an analog multiplexer. Once every frame, a strobe signal is also multiplexed onto the output data line. This strobe signal is used by the external electronics to regenerate the on-chip clock. The multiplexer is, in turn, driven by an 11-stage two-phase dynamic shift register, which operates from an on-chip two-phase clock. The clock frequency of 176 kHz is chosen so that a frame frequency of 16 kHz is obtained to be used by the test-waveform generation circuitry. The shift register is reset through the power-on-reset circuitry, which is activated when the power line is initially turned on. The output of the multiplexer is then passed through a unity gain broadband output buffer, which has a low output impedance and drives the output data lead. The on-chip electronics is also capable of self-testing the electrode impedance levels on demand.

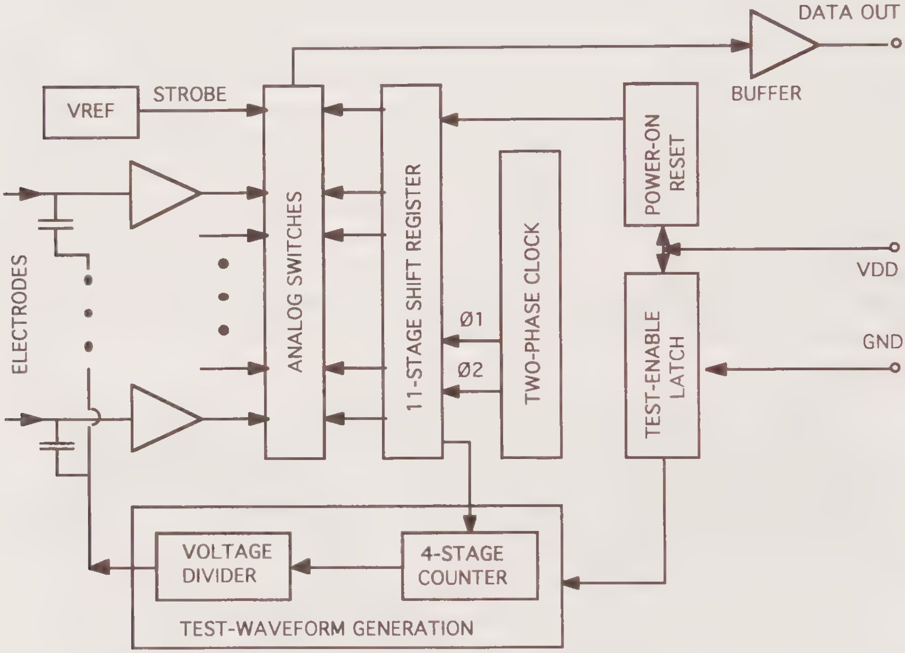

DATA OUT
BUFFER
VREF
STROBE
ELECTRODES
ANALOG SWITCHES
11-STAGE SHIFT REGISTER
Ø1
Ø2
TWO-PHASE CLOCK
POWER-ON RESET
VDD
GND
TEST-ENABLE LATCH
VOLTAGE DIVIDER
4-STAGE COUNTER
TEST-WAVEFORM GENERATION

FIG. 9.8 Detailed block diagram of a 10-electrode active recording microprobe with on-chip circuitry.[10]

When testing is desired, the power supply (V_{DD}) is pulsed from its normal value of 5 V to 8 V. This latches an on-chip test-enable latch in the TEST mode. After the circuit is latched in this mode, the supply is returned to its normal value. The test-enable latch then turns on the test-waveform generation circuitry. This circuitry receives a 16-kHz frame signal from the shift register and divides it down in frequency to 1 kHz through a 4-stage binary counter. The resultant 5-V, 1-kHz signal is then capacitively attenuated to an amplitude of 50 mV and is then applied capacitively to all the electrodes using 60 fF on-chip coupling capacitors. The amplitudes of the resulting induced electrode voltages are then amplified and multiplexed out in the normal way, providing an indication of the electrode impedance levels and a test of the functionality of the on-chip electronics. After testing, the circuitry can be returned back to its normal mode by turning the power supply off and then back on again.

The above design for the on-chip electronics requires only three output leads (for power, ground, and data) while maintaining the on-chip self-test capability. The circuitry has been designed in both 6μm enhancement-depletion n-channel metal-oxide-silicon (ED-NMOS) technology and in a 3-μm bulk p-well complementary metal-oxide-silicon (CMOS) technology. Figure 9.9 shows an optical photograph of the active probe that is passed through the eye of a needle. The probe is 15 μm thick, contains ten gold recording sites with 100-μm center-to-center spacing and each with an area of 64 μm^2, and has a shank that is 3.2 mm long and tapers from about 10 μm close to the tip to about 160 μm at the bottom. The circuitry occupies an area of 1.3 mm^2, and consumes about 5 mW of power from a single 5-V power supply. The details of the process integration and fabrication of these active probes are presented in Ref. 10. These active probes have been used in recording experiments.

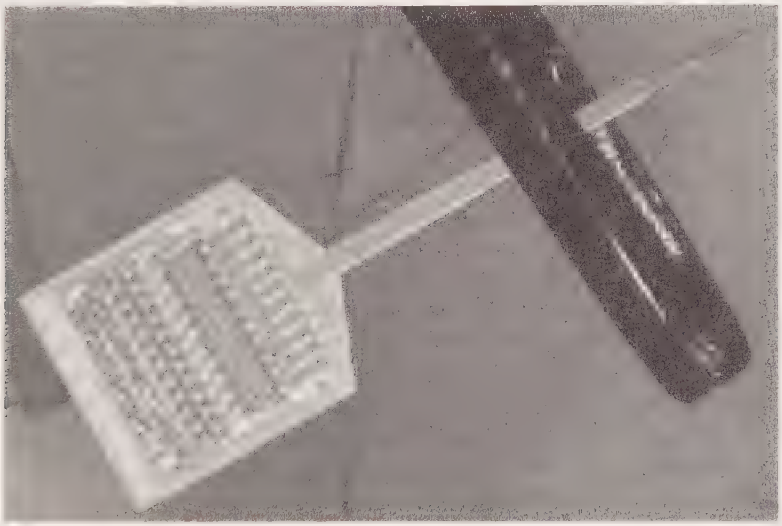


FIG. 9.9 Photograph of an active probe passed through the eye of a needle. The probe is 15 μm thick, contains ten gold recording sites with 100 μm center-to-center spacing and each with an area of 64 μm^2, and has a shank that is 3.2 mm long and tapers from about 10 μm close to the tip to about 160 μm at the bottom.

Although the 10-electrode recording microprobe described above is useful in many recording applications, it is desirable to enhance its functionality by increasing the number of available electrodes to record from, and by developing a two-dimensional probe that can be programmed by the user for improved site positioning. A 32-electrode active recording microprobe has also been developed.[27] This 32-electrode active probe preserves all of the features provided by the 10-electrode design, including per-channel amplification, analog multiplexing for reducing the output lead count, low output impedance, self-test capability, use of only three output leads, single power supply operation, low power dissipation, and small area. In addition, this probe allows the recording sites to be electronically positioned (by selecting eight of the 32 sites for amplification), incorporates dc feedback in the preamplifiers to eliminate the effects of dc potential variations at the recording sites and dc amplifier offsets, and utilizes one lead for both data input and output.

The on-chip circuitry consists of eight major circuit blocks: the front-end selector, preamplifier, analog multiplexer, clock, self-test, output buffer, I/O control, and input data/clock separator.[27] The 32 electrodes, placed on the micromachined silicon probe shank, are connected to the front-end electrode selector, a 32-to-8 switching array, where eight of the 32 signals are selected. The eight selected signals are then fed into the eight preamplifiers, each having a gain of 150–300 and a 3-dB signal bandwidth from 100 Hz to 10 kHz. The amplified signals are then sampled, multiplexed, and buffered out over the single I/O lead (when the I/O line is set in the output mode). The multiplexer is controlled by a two-phase non-overlapping clock generated on-chip. The clock signal is sent out together with the multiplexed neural signals in the form of frame marks, which are used by the external electronics to regenerate the clock. The regenerated clock is then used to control the demultiplexer to separate and reconstruct the eight individual signals. The recovered signals are then in analog format again and ready for further processing. The external electronics also performs per-channel spike counting and sends the results through the computer interface to the host computer where the results get displayed on the screen. The spike counting results give the user a rough but useful idea of how the selected recording sites couple with active neurons. If the user wants to change the configuration of the sites, he/she can enter the numbers of the desired sites from the keyboard, and then the computer sends that information to the external electronics. In turn, the external electronics first sends a signal to switch the on-chip I/O controller to the input mode and then superimposes the control data from the computer onto the synchronization clock and sends them over the same I/O line to the probe on-chip data/clock separator. The data/clock separator separates the control data and the synchronization clock and sends them to the front-end electrode selector. When all the input control data have been shifted into the front-end selector, the I/O line is switched back to the output mode and recording resumes with a new set of recording sites. The I/O line is, therefore, shared by both input and output functions. The power supply line is used not only to supply power but also to control the self-test signal generator.

The front-end electrode selector selects eight electrodes from the 32 that are available and connects them to the eight preamplifiers. The use of such a selector is a tradeoff; many recording sites are required for neurophysiological studies, but not so many can be accommodated with limited power dissipation and chip area. Also, since not all the sites couple to active neurons all the time, with a front-end electrode selector, the limited signal processing resources (both on- and off-chip) can be matched to those sites that are close to active neurons. In addition, the selector can be used to move the sites electronically if the relative position of the probe changes in chronic implant situations.

The on-chip circuitry designed for the 32-electrode multiplexed active recording probe uses only three I/O leads to perform normal recording, to enable self test, and to input control data and clock. The front-end selector makes it possible to electronically configure the recording site positions. The amplifier design achieves a unity gain for input dc drift and offset signals and a gain of 300 for signals from 100 Hz to 10 kHz; therefore, the input dynamic range of this amplifier is 1 mV ac on 100 mV of dc noise. The circuitry is designed in a p-well 3-μm single-metal double-poly bulk CMOS process, occupies an area of 2.5 mm^2, and dissipates 2.5 mW of power from a single power supply.[27]

The fabrication process for the active probes will not be discussed here. The reader is referred to other papers that have been published.[28] Other investigators have also developed different fabrication technologies to produce multielectrode probes for recording from the central nervous system.[29] Although the fabrication technology for these probes is different than that presented here, it still uses integrated circuit and silicon micromachining techniques to achieve small size, improved functionality, and long-term reliability. In either case, these multichannel recording probes are now in widespread use by many neurophysiologists and have produced important information on the signal processing capabilities of the central nervous system. It is believed that as they are further developed and as their operation in biological tissue is better understood, they will be employed in implantable systems for applications in auditory and visual prostheses.

9.3.2 Nerve Regeneration Electrodes

Recording microprobes presented above are employed as penetrating probes for recording neural signals from the central nervous system. These microprobes record signals generated and processed by the various regions of the brain and before they are transmitted to organs located throughout the body through the peripheral nervous system. In many applications, it is more desirable to record neural activity that is traveling through the nerve fibers because these signals are more locally targeted and contain more specific information and require much less signal processing. To achieve this goal several investigators have suggested and studied the use of a sieve electrode array in which a thin diaphragm with many small holes is positioned between the cut ends of a peripheral nerve;[30,31] the nerve is left to regenerate through the holes and reinnervate its target organ, as illustrated in Fig. 9.10. Since the positions of the regenerated axons relative to recording/stimulating sites fabricated adjacent to the holes are fixed, long-term recording and stimulation of axons can be successfully achieved. The recordings obtained this way can help to further our understanding of the nervous system and can be used to generate control signals for manipulation of prosthetic devices for amputees and for stimulation of paralyzed muscles. A thorough review of the history of the development of sieve electrodes has recently been presented by Kovacs et al.[30] Akin et al. have fabricated sieve electrodes for nerve regeneration applications using the above technology.[31] Figure 9.11 shows a scanning electron microscope (SEM) view of one of these sieve electrodes. The electrode consists of a thin diaphragm in the middle with a diameter of 400 μm. The diaphragm contains many through via slits to allow regeneration of the nerve through it, and also supports recording sites that are placed near the slits to record electrical signals traveling through the nerve. The slits are about 5 μm wide and 300 μm long.

The fabrication process for these electrodes is very similar to that used for making the penetrating probes discussed above.[31] The silicon ribbon cable, which provides interconnection between the sites and an external connector located outside, is also fabricated using the same process used for the probe. Here, polysilicon conductors are supported on a 4-μm-thick silicon cable, and are insulated above and below by deposited dielectric films.[32]

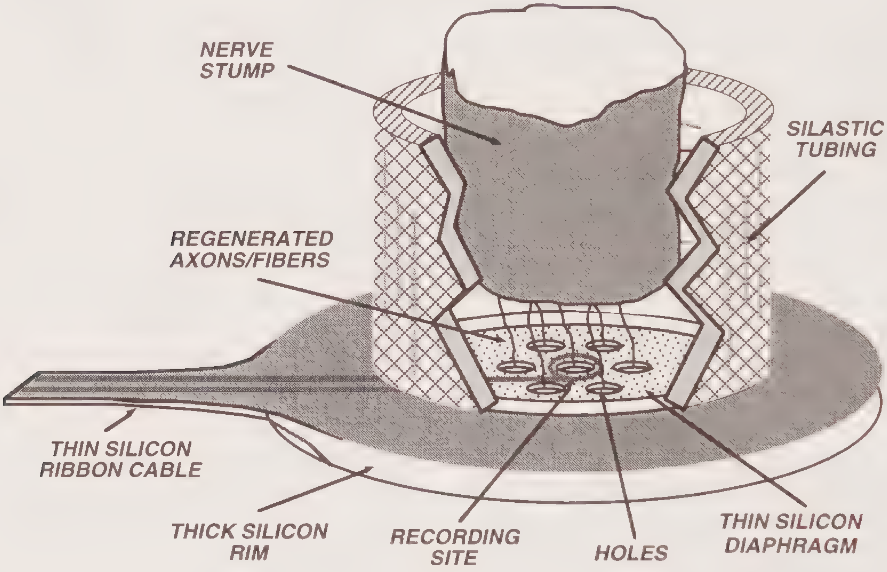

NERVE STUMP
SILASTIC TUBING
REGENERATED AXONS/FIBERS
THIN SILICON RIBBON CABLE
THICK SILICON RIM
RECORDING SITE
HOLES
THIN SILICON DIAPHRAGM

FIG. 9.10 Structure of a micromachined silicon sieve electrode for nerve regeneration applications.[31]

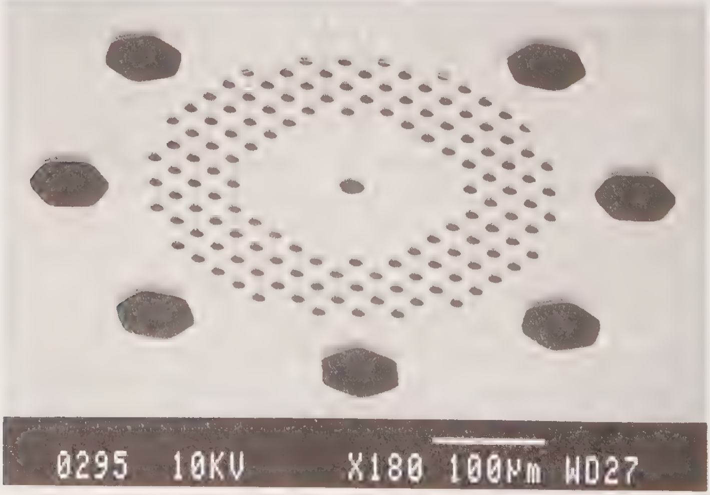

0295 10KV X180 100μm WD27

FIG. 9.11 Scanning electron microscope (SEM) view of a micromachined silicon sieve electrode.[31]

An important feature of this peripheral recording system is the manner in which it acquires and delivers recorded neural signals to the outside world. As shown in Fig. 9.10, this implantable system contains electronic circuitry for amplifying, digitizing, and transmitting neural signals to the outside world using a fully integrated radio-frequency (RF) telemetry system.[33] The system is capable of performing signal amplification and filtering, low-power analog-digital conversion, bi-directional user interface, and RF telemetry for power/data transfer. Telemetry systems are a very important and critical component in many future implantable biomedical systems because they eliminate the need for hard-wired connection between the implant and the outside world, which causes many problems with long-term stability, infection, discomfort to the patient, and frequent surgery to replace implanted batteries. Because of their importance, we will describe the design of the telemetry circuitry used in these sieve electrodes in more detail below.

Figure 9.12 shows the block diagram for the telemetry circuitry along with the communication protocol for the system. The communication protocol supports two modes of operation: write address and read data. The serial digital signals 1 and 0 are encoded by long and short pulse durations, respectively. Following the start signal, a write command (digital 1) is sent, followed by the 10-bit address (5 bits per channel) of the two recording channels that can be selected from among a set of 32 recording channels. Once the channel selection information is transmitted to the on-chip electronics, it waits for the read command (digital 0) to start analog-to-digital conversion of the two channels consecutively and transmits back a total of 16 data bits (8 bits per channel). Then, the on-chip circuitry waits for the next command, and depending on the mode, it either selects the new channels to be recorded from or starts another conversion and transmission cycle. This protocol allows easy synchronization, gives the user control over sampling rate, and is compatible with our RF telemetry system that uses amplitude modulation (AM) and passive reverse telemetry for bi-directional signal transmission. The implementation of the RF telemetry interface and mixed-mode signal processing electronics is achieved using a BiCMOS technology, which provides high performance, junction-isolated vertical bipolar transistors and junction and Zener diodes. This system has four major circuit blocks for voltage regulation, data detection, amplification, and analog-digital conversion.

9.3.2.1 Voltage regulator

The 5-V supply voltage necessary for the implanted electronics is generated by the voltage regulator circuit shown in Fig. 9.13. The RF signal induced across the receiver coil, which is about 20-V peak, is first rectified by the full-wave rectifier using diodes D1–D8. Good voltage regulation is achieved by setting a constant current in diodes D16-25 through transistors M8, M9 and diodes D13–15. Darlington transistors Q1 and Q2 supply the current to the circuit. Transistors M10-13 and diodes D26–29 keep the voltage regulated against load current variations. In order to improve voltage regulation (necessary for the accuracy of the ADC) and to save lateral area of the implantable unit, capacitors C1, C2, and

C3 are chosen to be hybrid capacitors. Simulation and measured results show that the maximum peak-to-peak ripple on top of 5-V supply is <0.02 V (better than 8 bits). This regulation is maintained during the loading of the receiver coil for reverse passive telemetry using the circuit made of transistors M1–7 and diodes D9–12.

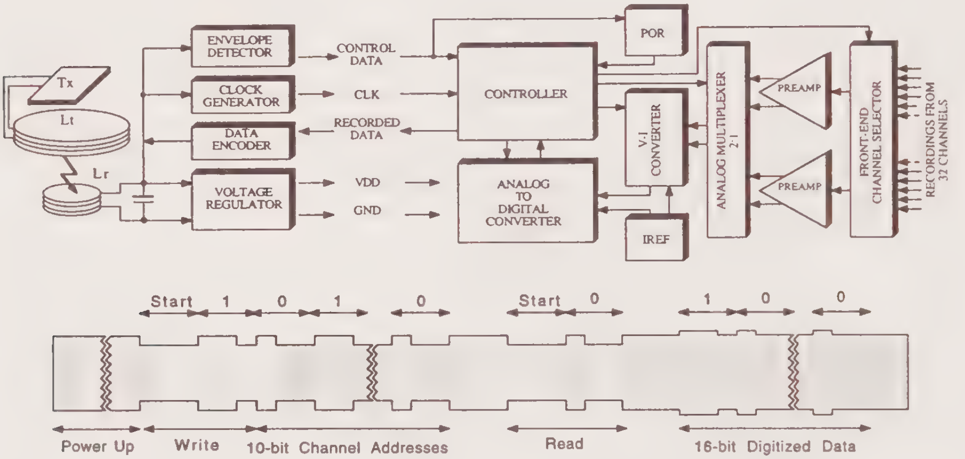

ENVELOPE DETECTOR
CLOCK GENERATOR
DATA ENCODER
VOLTAGE REGULATOR
CONTROL DATA
CLK
RECORDED DATA
VDD
GND
Tx
Lt
Lr
CONTROLLER
ANALOG TO DIGITAL CONVERTER
POR
V-I CONVERTER
IREF
ANALOG MULTIPLEXER 2:1
PREAMP
PREAMP
FRONT-END CHANNEL SELECTOR
RECORDINGS FROM 32 CHANNELS
Start
1
0
1
0
Start
0
1
0
0
Power Up
Write
10-bit Channel Addresses
Read
16-bit Digitized Data

FIG. 9.12 Block diagram of the on-chip circuitry, along with the communication protocol for the system.[33]

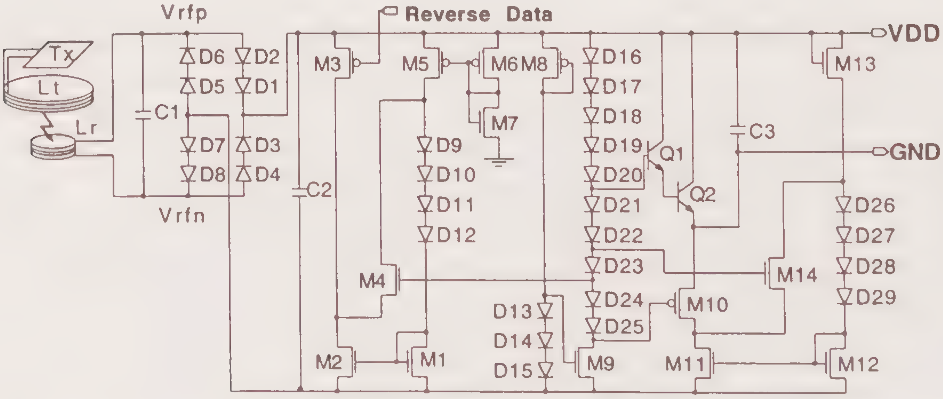
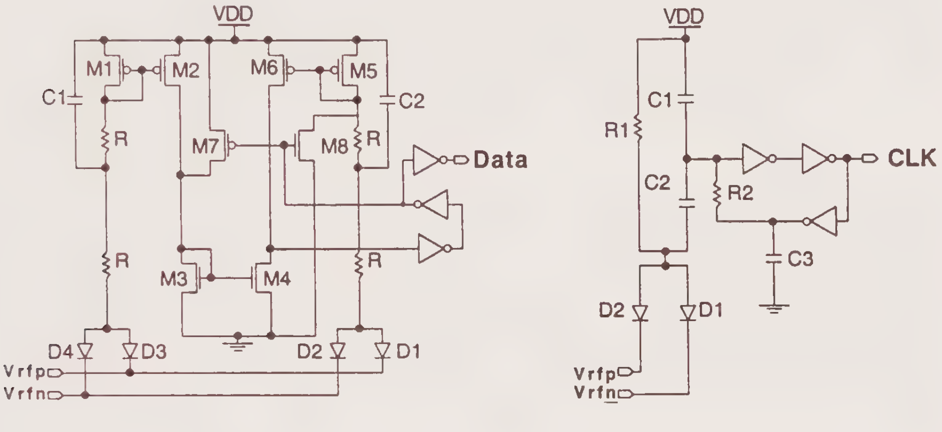

Vrfp
Vrfn
Reverse Data
VDD
GND
Tx
Lt
Lr
C1
C2
C3
D1
D2
D3
D4
D5
D6
D7
D8
D9
D10
D11
D12
D13
D14
D15
D16
D17
D18
D19
D20
D21
D22
D23
D24
D25
D26
D27
D28
D29
M1
M2
M3
M4
M5
M6
M7
M8
M9
M10
M11
M12
M13
M14
Q1
Q2

FIG. 9.13 The voltage regulator for the 5V supply shown with the circuitry used for reverse passive telemetry.[33]

9.3.2.2 Data detector

Data detection is accomplished using a new current-mode envelope detector, shown in Figure 9.14(a). In this circuit, the current through transistors M1 and M5 is different under transient conditions due to different capacitor values for C1 and C2. Therefore, any amplitude change in the RF signal (Vrf) will result in a change in currents through M1 and M5, which are mirrored to M4 and M6, so that they can be compared. Transistors M7 and M8 are used to generate hysteresis to prevent any state change due to ripples on the currents. The detector occupies an area of 0.5 mm^2 (single-metal 3-μm CMOS), and consumes a power of only 2 mW. Figure 9.14(b) also shows the circuit used for extracting the clock signal from the RF carrier.



(a) (b)

FIG. 9.14 The current-mode envelop detector.[33]

9.3.2.3 Preamplifiers

±500-μV neural signals are amplified with two preamplifiers as shown in Fig. 9.15. Each preamplifier has an in-band ac gain of 40 dB with a 3-dB bandwidth of 100 Hz to 3.1 kHz. A special feature of the preamplifier is that it has a dc gain of -40 dB, which is critical in preventing amplifier saturation due to input offset and dc drift caused by the electrode-electrolyte half-cell. The low cutoff frequency required for dc suppression is achieved using a diode-capacitor filter, which prevents the use of chopping or off-chip hybrid components and allows the preamplifier to accommodate a ±500-μV signal superimposed on ±160-mV dc or low-frequency noise. Each amplifier dissipates 0.3 mW, occupies 0.2 mm^2, generates the differential signal for the voltage-to-current converter, and incorporates feedback to prevent output saturation even in the face of very gross process variations.

9.3.2.4 Analog-to-digital converter

A current-mode algorithmic AD conversion approach is used for data conversion, as shown in Fig. 9.16. Conversion is achieved algorithmically by multiplying the input current (Iin) by two and then comparing it with the reference current (Iref). If 2 Iin is larger than Iref, then the bit output is "1" and the current, 2 Iin-Iref, is fed to the circuit to determine the next bit. Otherwise the bit output is "0", and 2 Iin is fed into the circuit to determine the next bit. This continues until the required bits are achieved. The reference current is set at 102.4 μA, i.e., a maximum current of only 204.8 μA circulates in the circuit at a time. The converter combines the amplifier and the current comparator circuitry and consumes only 2mW, while achieving a conversion accuracy of better than 8 bits with a conversion time of 8 μsec per bit. The analog portion of the converter occupies only 0.13 mm^2, while the digital portion occupies 1.1 mm^2 in a 3-μm CMOS process. Note that the low power dissipation of this converter together with its very fast conversion rate are desirable for the system being proposed here. We will use either this converter or a successive approximation converter for the subdural platform being proposed here.

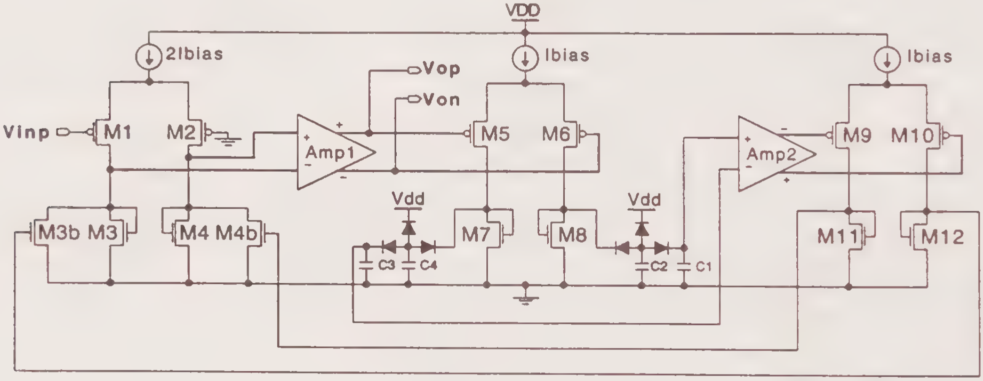

VDD
2Ibias
Ibias
Ibias
Vop
Von
Vinp
M1
M2
Amp1
M5
M6
Amp2
M9
M10
Vdd
Vdd
M3b M3
M4 M4b
M7
M8
C3
C4
C2
C1
M11
M12

FIG. 9.15 A simplified schematic of the preamplifier.[33]

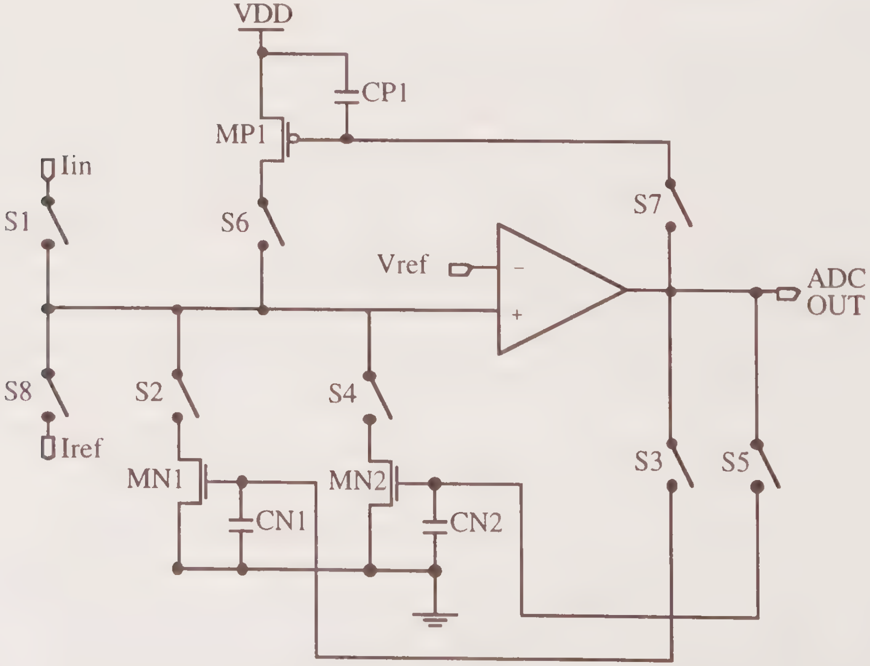

VDD
CP1
MP1
Iin
S1
S6
S7
Vref
ADC OUT
S8
S2
S4
Iref
MN1
MN2
CN1
CN2
S3
S5

FIG. 9.16 The current-mode algorithmic analog-to-digital converter.[33]

All the necessary power supply voltages and reference signals for the converter are derived on-chip. The full functionality of the on-chip circuitry described above has been verified. The fabricated chip, together with the hybrid capacitors and receiver coil, and a nerve regeneration electrode are shown in Fig. 9.17. The circuit chip measures 4×6 mm^2, contains more than 5000 transistors, and dissipates ~90 mW of power. The majority (80 mW) of this power is consumed by the 5-V voltage regulator that provides better than 99.8% regulation (needed for the 8-bit converter). Figure 9.18 shows a ±400-μV signal fed into the preamplifier and the digitized output signal recovered by the external electronics. The voltage signal is amplified, multiplexed, and converted to a current signal, and is then passed through the analog-digital converter before transmission to the outside world. The size of the total implantable unit, including three hybrid capacitors and the receiver coil, is 5×8×2 mm^3.

In addition to recording electrical activity from biological systems, it is often required that electrical stimuli are delivered to the tissue through appropriate electrodes. Electrical stimulation is needed for prosthetic applications ranging from stimulation of the neural tissue for restoration of functions like hearing and vision, to the stimulation of muscular tissue for activating various muscles that have lost connection to the nervous systems. In the following we briefly review two stimulating systems. The first is a multichannel stimulating probe for neural applications, while the second is an implantable microstimulator for functional neuromuscular stimulation applications.

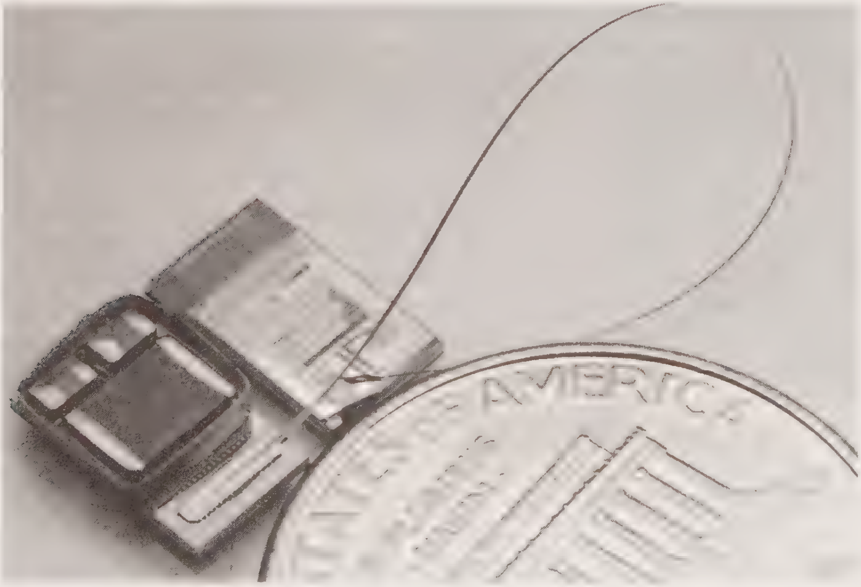


FIG. 9.17 Photograph of a prototype telemetry system showing the RF receiver coil and storage capacitors on a complementary metal oxide semiconductor (CMOS) circuit chip. The size of the total implantable unit, including 3 hybrid capacitors and the receiver coil, is 5×8×2 mm^3.[33]

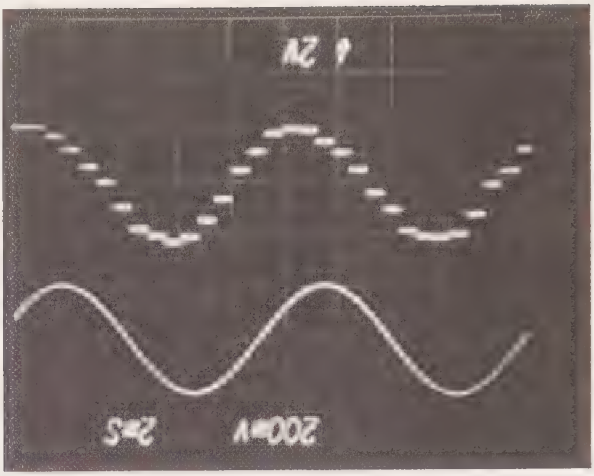


FIG. 9.18 A ±400-μV signal fed into the preamplifier and the digitized output signal recovered by the external electronics. The voltage signal is amplified, multiplexed, and converted to a current signal, and is passed through the analog-digital converter before transmission to the outside world.[33]

9.3.3 Stimulating Microprobes

Neural electrical stimulation is usually achieved by using microelectrodes to deliver small currents into the neural tissue. Stimulating microprobes for many prosthetic applications should satisfy a number of requirements. They should be able to deliver current into tissue through multiple sites precisely spaced on a substrate that is small in size to minimize damage to the surrounding tissue. The stimulating sites should be made out of a material that is biocompatible and that does not degrade in terms of their charge delivery capabilities after years of operation in the biological tissue. In addition, the stimulating sites should possess a high charge delivery capacity so that large amounts of charge can be delivered through small sites. One of the most important requirements of a multichannel stimulating electrode is that a minimum number of leads be connected to it. This is required for improved long-term reliability of the implant, for reducing the tethering forces exerted on the implanted probe due to the interconnect leads, and for simplifying packaging and assembly. In order to achieve these requirements, active multichannel stimulating probes with on-chip multiplexing and current drivers have been developed.[34]

The overall structure of active stimulating microprobes is the same as recording microprobes shown in Fig. 9.3. The probe is implemented using silicon micromachining to form probe shanks that support an array of stimulating sites, and houses multiplexing and drive electronics on its back end, fabricated using CMOS technology. The block diagram of a 16-channel active stimulating microprobe is shown in Fig. 9.19.[34] This device has a digital-to-analog-converter (DAC) for every channel, allowing all sixteen channels to be actively stimulating at any given time, with individual control over the current amplitudes. The circuit

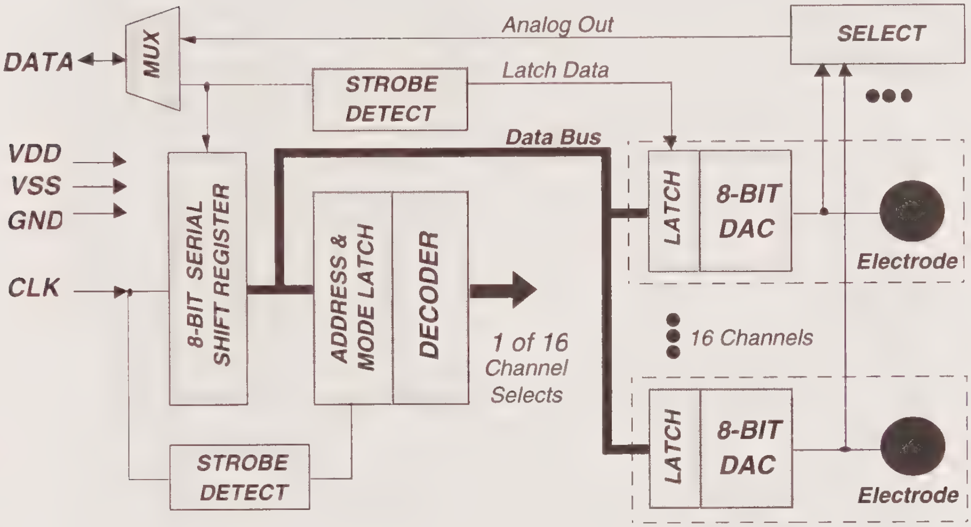

DATA
MUX
VDD
VSS
GND
CLK
Analog Out
SELECT
STROBE DETECT
Latch Data
Data Bus
8-BIT SERIAL SHIFT REGISTER
ADDRESS & MODE LATCH
DECODER
1 of 16 Channel Selects
LATCH
8-BIT DAC
Electrode
16 Channels
LATCH
8-BIT DAC
Electrode
STROBE DETECT

FIG. 9.19 Block diagram of a 16-electrode stimulating microprobe for neural stimulation.[34]

receives 16-bit data words that are shifted into the data pad using a 4-MHz clock. Four bits represent the site address, two are mode bits, eight set the current amplitude, and two are used for strobing and status. In the normal mode of operation, the specified current is set up on the addressed channel and remains on until the channel is accessed again to alter it. Other channels are set up sequentially to source or sink current as desired. At a clock rate of 4 MHz, the time resolution between any two changes in the probe current pattern is 4 μs, which is effectively simultaneous to the tissue. A second mode of operation grounds the associated stimulus channel, allowing the DAC currents to be calibrated externally without stimulating the tissue. A third mode of operation allows the analog voltage developed in response to stimulus current to be observed externally to check site integrity at any time. Finally, a fourth mode of operation biases the electrode above ground between pulses to increase the charge delivery of the iridium oxide electrode sites.

Figure 9.20 shows a photograph of a fabricated stimulating microprobe. This probe consists of two shanks, each of which supports 8 stimulating sites. Each shank is 3 mm long and 135 μm wide, and the shanks are 500 μm apart. The site areas are 1000 μm^2 and are spaced 250 μm apart. The circuitry is fabricated using a 3-μm p-well single-metal single-poly bulk CMOS process, occupies an area of 4.4 mm × 2.5 mm, contains about 7100 transistors, consumes 80 μW of power when idle, operates from ±5-V supplies, and operates at clock rates as high as 10 MHz. The output currents are linear to 8 bits with sourcing and sinking currents matched within ±1/2 LSB, ensuring adequate charge balancing at the stimulating sites.[34]

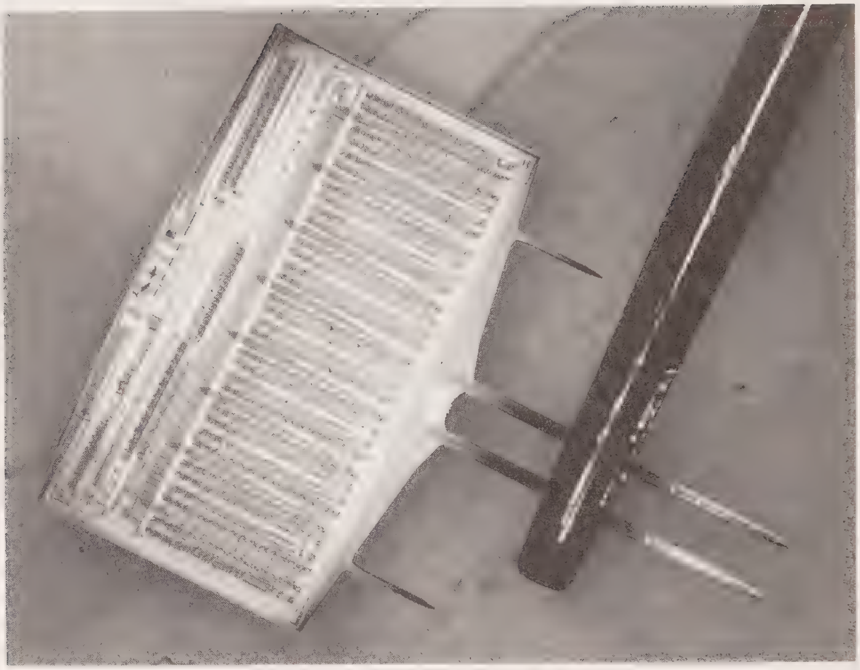


FIG. 9.20 Photomicrograph of the 16-channel stimulating microprobe. Each shank supports 8 stimulating sites and is 3mm long and 135μm wide.[34]

In addition to the difference in the function of the circuitry between the recording and stimulating probes, the stimulating sites are also made from a different material in order to satisfy the requirements needed in stimulation applications. The stimulating sites are required to meet very heavy demands for chronic stimulation applications. They must be able to withstand long-term exposure to body fluids and should be capable of delivering large amounts of charge without dissolution into the body or delamination from the conductor surface. Furthermore, the charge density must not exceed the gas evolution threshold at the stimulating site. A material that has shown promise in meeting the requirements for high charge capacity stimulating sites is iridium oxide. Iridium is the densest periodic element known, it is highly resistive to corrosion under static potential conditions, and is chemically inactive since it is a noble metal.[35] This material will grow a hydrous oxide layer on its surface when electrically activated in acidic or saline solutions.[35] The oxide exhibits high charge delivery capability that is believed to be a consequence of oxidation state changes that occur in response to varying voltage levels.[35]

Iridium sites can be activated through cyclic voltammetry,[35] by immersing the electrode in a phosphate buffered saline solution and cycling it between potential levels of -1 to 1.1 volts with respect to a SCE reference electrode using a triangular voltage ramp at 2.1 V/sec.[35] The sites are typically activated to obtain a charge capacity of ≈100 mC/cm^2. When such an electrode is used in stimulating applications where it is subjected to current waveforms that are only several hundred microseconds in length, it can only utilize a portion of its full charge capacity. For the microelectrodes reported here this charge delivery capability is measured to be about $4mC/cm^2$. This is still orders of magnitude larger than is possible to obtain with other materials such as stainless steel and platinum.[34,35] These sites have been subjected to long-term current pulse tests and have not shown any sign of degradation after several hundred million pulses.[34-36] The large charge delivery capability and the long-term stability of iridium oxide sites are important characteristics that are crucial to many applications.

9.3.4 Muscular Microstimulators

In addition to probes for neural stimulation, many applications require microstimulators for muscular stimulation. Functional neuromuscular stimulation (FNS) has been widely used to restore function in various applications, including stimulation of paralyzed muscles. FNS has been successfully demonstrated in stimulating paralyzed muscles of paraplegic and quadriplegic patients in many different clinical settings.[5,6] The majority of stimulators developed use pulses of electrical current to deliver controlled amounts of charge into the muscle. To date, most of these stimulators have typically been fabricated using hybrid elements for the receiver circuitry, long metal wires for stimulating electrodes, and large packages. Although these systems have demonstrated the benefits of FNS, it is believed that in order to fully realize these benefits, microstimulators that contain both electrodes and full circuitry in one small injectable device are required. In order to implement such a microstimulator, many new technologies have been developed, including power and data transfer for microdevices,[11] packaging technologies,[37] and techniques for fabrication of miniature stimulating electrodes and control electronics.[38]

In order to overcome the above shortcomings, an implantable, injectable microstimulator for functional neuromuscular stimulation has been developed.[38] The microstimulator can be injected into muscle by expulsion through a hypodermic needle. The overall structure of a single-channel implantable microstimulator is shown in Figure 9.21. It consists of a micromachined silicon substrate that incorporates two stimulating electrodes which are the cathode and anode of a bipolar stimulating electrode pair; a hybrid-connected tantalum chip capacitor for power storage; a receiving coil; a bipolar-CMOS integrated circuit chip for power regulation and control of the microstimulator; and a custom made glass capsule that is electrostatically bonded to the silicon carrier to provide a hermetic package for the receiver-stimulator circuitry and hybrid elements. The stimulating electrode pair resides outside of the package and feedthroughs are used to connect the internal electronics to the electrodes. The size of the microstimulator is constrained such that it can be implanted into muscle tissue through the use of a hypodermic needle. The overall dimensions of the implant are $2x2x10mm^3$. Additionally, power and data are transmitted using RF-telemetry, and the microstimulator is capable of delivering a constant current pulse of 10mA into loads of <800W for up to 200µsec durations and be able to repeat stimulation at a rate of up to 40Hz. The microstimulators are also addressable such that up to 32 stimulators can be driven using a single transmitter coil.

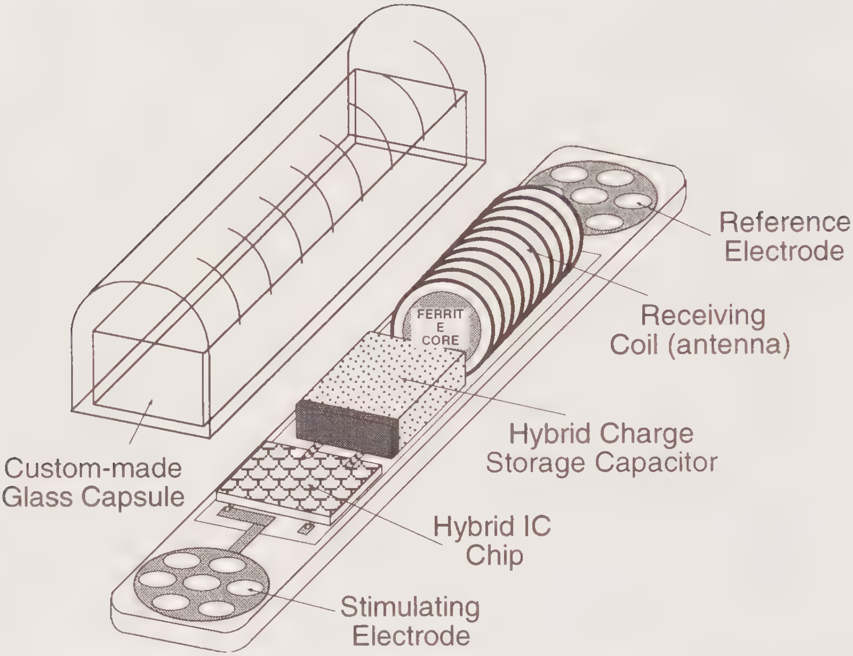


FIG. 9.21 Overall structure of a single-channel implantable microstimulator for functional neuromuscular stimulation.[38]

Figure 9.22 shows the overall system electronics required for the microstimulator, and the power and data transmission protocol used for radiofrequency telemetry. The circuit receives an amplitude modulated RF carrier from an external transmitter and generates 8-V and 4-V dc supplies, generates a clock from the carrier signal, decodes the modulated control data, interprets the control data, and generates a constant current output pulse when appropriate. The RF carrier used for the telemetry link has a nominal frequency of ≈1.8 MHz, and is amplitude modulated to encode control data. Logical "1" and "0" are encoded by varying the width of the amplitude modulated carrier, as shown in Figure 9.20. The carrier signal is initially high when the transmitter is turned on and sets up an electromagnetic field inside the transmitter coil. The energy in the field is picked up by receiver coils, and is used to charge the hybrid capacitor. After power-on, the carrier amplitude is momentarily turned low to reset all the circuit blocks on the receiver chip, and is then modulated to transmit a 5-bit address to 32 microstimulators located within the volume of the transmitter coil. All microstimulators check the transmitted address against that programmed into their circuitry; the one that matches its address to the transmitted address will continue monitoring the carrier for additional data. The rest of the microstimulators are reset and await the receipt of another set of control data. After a specific microstimulator is selected, the carrier signal is turned high and then back down again, and is maintained at the low level for a period between 1–200 μsec. The selected microstimulator will then deliver a constant current pulse into the tissue through the stimulating electrode pair for the period that the carrier is low. Finally, the carrier is turned back high again, which will indicate the end of the stimulation period to the selected microstimulator, thus allowing it to charge its capacitor back up to the on-chip voltage supply.

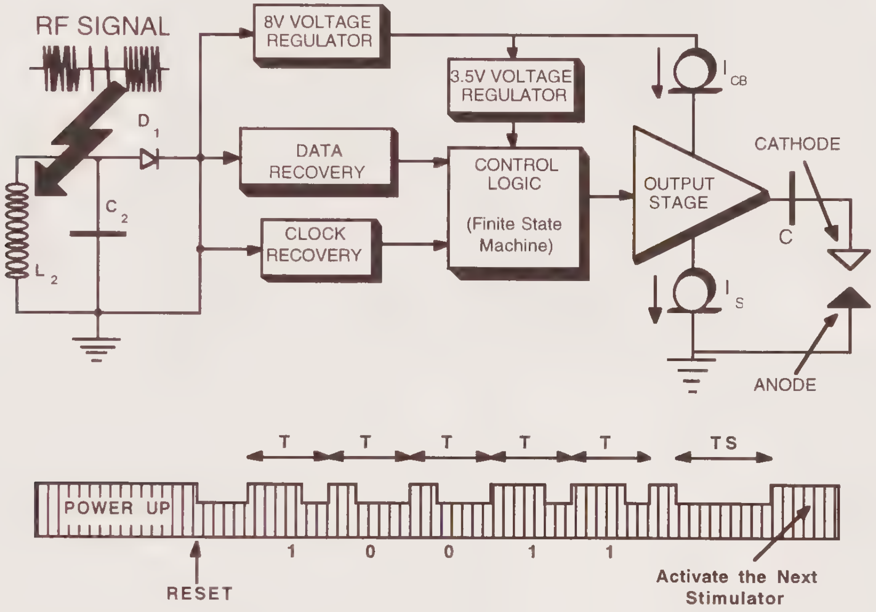

RF SIGNAL
8V VOLTAGE REGULATOR
3.5V VOLTAGE REGULATOR
I_{CB}
D_1
DATA RECOVERY
CONTROL LOGIC
(Finite State Machine)
OUTPUT STAGE
CATHODE
C_2
CLOCK RECOVERY
C
L_2
I_S
ANODE
T
T
T
T
T
TS
POWER UP
1
0
0
1
1
RESET
Activate the Next Stimulator

FIG. 9.22 Overall system electronics required for the microstimulator, and the power and data transmission protocol used for radiofrequency telemetry.[38]

A variety of passive and active electronic devices are used in the implementation of this circuitry, including MOS and bipolar transistors, pn junction diodes, Zener diodes, resistors, and capacitors. In previous implantable stimulators, many of these elements were hybrid components that took up large amounts of space. For this microstimulator, the hybrid chip capacitor and receiving coil consume most of the available volume, so the rest of the stimulator circuitry must be made very small and without the use of hybrid components. The circuitry is implemented using a BiCMOS circuit fabrication technology.[36,38] The process combines a deep boron diffusion with a standard 3-µm p-well CMOS technology to implement a variety of active devices. The deep boron diffusion penetrates through a lightly doped n-epi region to connect with a p-substrate. This provides junction isolation of one n-epi region from another. These regions can be used to contain CMOS devices along with vertical or lateral npn bipolar transistors, lateral pnp bipolar transistors, p-well/n-epi diodes, p-well/n^+ diodes, or p^+/n^+ Zener diodes. In addition, a double polysilicon process allows for on-chip capacitors, and large resistances can be reliably obtained through the use of the various diffused regions.

On-chip circuitry has been designed to generate two regulated power supply voltages (4 V and 8 V) from the RF carrier, to demodulate the RF carrier in order to recover the control data that is used to program the microstimulator, to generate the clock used by the on-chip control circuitry, to deliver a constant current through a controlled current driver into the tissue, and to control the operation of the overall circuitry using a low-power CMOS logic controller. Figure 9.23 shows a complete microstimulator in the eye of a needle on a US penny for size comparisons. The hybrid chip capacitor, the coil, and the glass capsule used for hermetic packaging are all shown. Note that the glass capsule is removed to reveal the electronic portion of the microstimulator. The on-chip stimulating and return electrodes are located at the ends of the silicon substrate. Figure 9.24 shows the operation of the circuitry. Note that the bottom trace shows the modulated telemetry signal transmitted to the implant, the middle trace shows the demodulated control data received by the implant, and the top trace shows the current pulse delivered to the tissue through the stimulating electrode. This device is now being implanted in biological tissues for complete testing.

The examples presented above demonstrate clearly the advantages of micromachined sensors and actuators for use in implantable biomedical devices for prosthetic applications, as well as the critical role of signal processing and telemetry circuitry in these systems. As mentioned before, to make a complete system that can satisfy all of the requirements presented in the previous section, a number of technologies have to be employed. Micromachining drastically reduces the dimensions of implantable devices, while integrated circuitry significantly enhances the functionality and reliability of these devices. An important component in all of these implantable devices is the package, which ultimately determines the reliability, long-term stability, and biocompatibility of implantable biomedical systems. A hermetic packaging technology for use in implantable systems is presented in the following section.

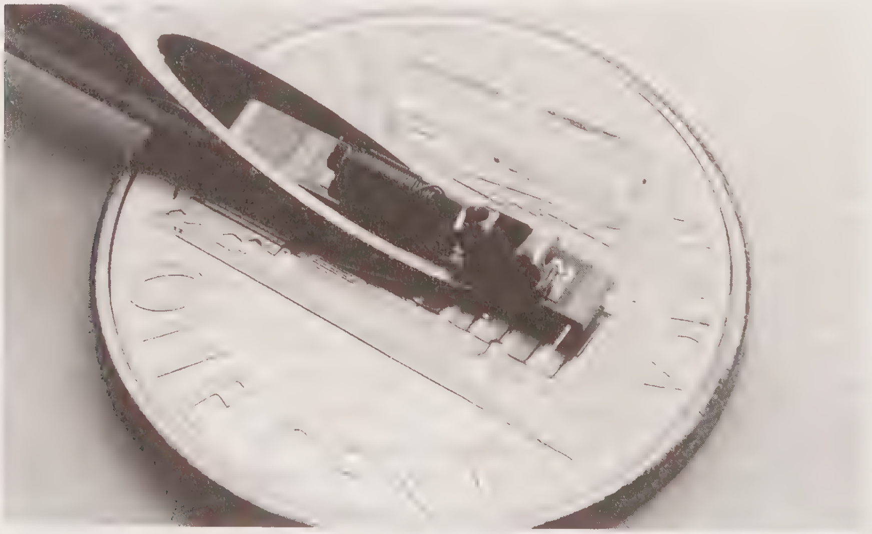


FIG. 9.23 A complete microstimulator in the eye of a needle on a US penny for size comparisons. The hybrid chip capacitor, the coil, and the glass capsule used for hermetic packaging are all shown.[38]

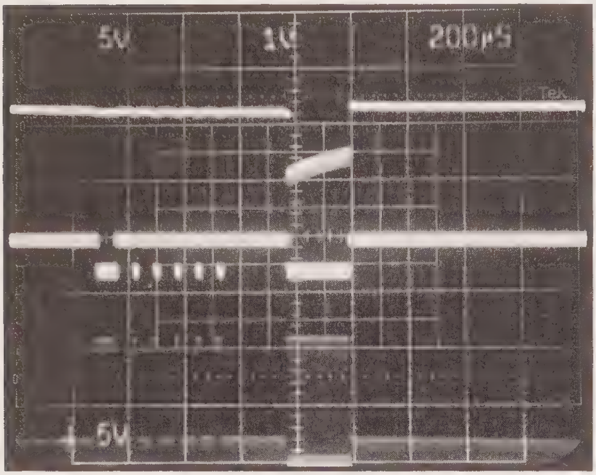


FIG. 9.24 Output of the microstimulator circuitry. The bottom trace shows the modulated telemetry signal transmitted to the implant, the middle trace shows the demodulated control data received by the implant, and the top trace shows the current pulse delivered to the tissue through the stimulating electrode.[38]

9.4 PACKAGING TECHNOLOGIES

Encapsulation and packaging of implantable biomedical systems for long-term implantation in the body is of paramount importance since most of these systems usually fail due to attack by biological fluids. Any package used in these systems should not only protect the transducers and their associated circuitry from the body fluids, but should also allow for contact between the sensing/actuating sites and the tissue. The requirement for this selective packaging has posed many challenges for implantable systems. In some applications, adding circuitry to the implantable module can simplify packaging. This may seem unjustified at first because adding circuitry to any transducer demands that this circuitry be protected from the outside environment. However, it should also be noted that the addition of the circuitry often solves more important problems and challenges. As was mentioned in the previous section, two of the important circuit functions are multiplexing for lead reduction and buffering for added drive capability and lower impedances. Both of these functions significantly simplify the packaging of implantable systems. Leads are often the most damage prone portion of any implantable biomedical device. It is therefore important that we not only reduce their numbers in order to simplify the packaging task and reduce any adverse mechanical trauma that too many leads can induce in the tissue, but also ensure that they are sufficiently buffered so that the electrical system remains operational in case the insulating materials forming the package are partially attacked by biological fluids. Therefore, adding electronics to implantable sensors improves both the mechanical and the electrical performance of the system.

Obviously, the circuitry itself needs to be protected from the biological environment. The degree of difficulty with which the on-chip circuitry can be encapsulated depends strongly on the structure of the transducers and their fabrication process. Hermetic as well as non-hermetic techniques have been and are being developed for encapsulating a variety of these systems.[39] Inorganic thin films such as silicon nitride and silicon dioxide, and organic thin film polymers such as polyimides, silicone, and Parylene have been used to encapsulate both sensors and electronics.[39-44] Hermetic packaging using silicon-glass electrostatic bonding and silicon-silicon bonding is becoming more feasible as these technologies mature and as new techniques for low-temperature bonding of glass to silicon are being developed.[45-51] In summary, protection of sensors as well as on-chip circuitry for most implantable systems is not an impossible task, and very often the advantages of having on-chip circuitry with regard to packaging outweighs the problems and disadvantages. In the following we review hermetic packaging techniques for implantable biomedical sensors.

9.4.1 Hermetic Packaging Techniques

Packaging has long been, and continues to be, a major source of failure in many implantable systems, and effective techniques need to be developed before implantable systems can be practically put to use. The pacemaker industry has to a large extent overcome this problem by using hermetically sealed metal cans formed from titanium. However, in many emerging biological applications a metal package cannot be used since these system utilize RF telemetry; a metal package can shield the antenna from the electromagnetic fields used for power and data transfer. In addition to this requirement, many new systems are much smaller than a standard pacemaker and it is difficult to utilize a miniature metal package. In order to satisfy these requirements, packages made from either silicon or glass are very attractive. Both silicon and some types of glass are biocompatible, can be micromachined to exact dimensions, and can be bonded to a number of different substrates using different bonding technologies.

In addition to the package, it is also required that electrical connections be created between the sensors residing outside of the package and the circuitry and other electronic elements protected inside the package. This requires that feedthroughs be fabricated through the package. These feedthroughs should be small, should not compromise the hermeticity of the package, and should be generally low-resistance.

In the following, a glass-silicon hermetic package that has been developed for the implantable microstimulator presented previously is described in detail. The technology for the formation of this package, and its feedthroughs, can be easily extended to other applications and other devices. Packages for most of these applications should provide a number of features, including:

- Protect a variety of electronics components from harsh body fluids for extended periods of time (many decades);

- Provide many feedthroughs between the sealed environment inside the package to external sensors and stimulators in a small area;
- Support a number of different device configurations and structures with varying number of feedthroughs and dimensions;
- Be biocompatible and not cause adverse long-term reactions in the body.

To achieve these goals, we have developed a glass-silicon package, as illustrated in Fig. 9.25.[37] The packaging technique uses low temperature (320°C) electrostatic bonding of a custom-made glass capsule (Corning # 7740, $2\times2\times8$ mm^3) to fine grain polysilicon in order to form a hermetically sealed cavity. Electrostatic (or anodic) bonding technique can be used for permanently bonding silicon to glass.[51] In this process, a silicon wafer is placed against a glass wafer that has a similar thermal expansion coefficient as silicon. Corning type 7740 glass with a thermal expansion coefficient of $\sim3.2\times10^{-6}$/°C is widely used for this purpose (silicon's is 2.6×10^{-6}/°C). The glass-silicon sandwich is heated to 300–450°C, and a voltage of several hundred volts (~200 V to 2000 V) is applied across the wafers with the glass biased negatively. At these temperatures, glass is a weak conductor, so most of the applied voltage is dropped across the interface between the glass and the silicon. The high field in that region pulls the two wafers into intimate contact and brings oxygen out of the glass to bond with the silicon surface. The nature of the bond is chemical and is due to the formation of a thin silicon dioxide layer at the interface. The bond is stronger than either of the two materials (i.e., silicon and glass), and any attempt to break the bond would result in breaking either glass or silicon. The quality of the bond depends on the surface roughness and cleanness of the silicon and glass in the areas where they

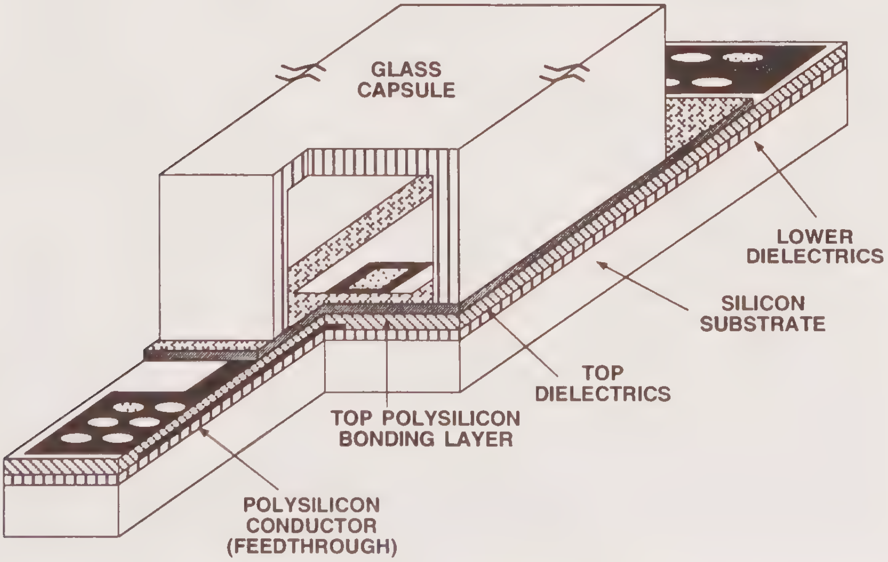

GLASS CAPSULE
LOWER DIELECTRICS
SILICON SUBSTRATE
TOP DIELECTRICS
TOP POLYSILICON BONDING LAYER
POLYSILICON CONDUCTOR (FEEDTHROUGH)

FIG. 9.25 The structure of a hermetic glass-silicon package for biomedical systems.[37]

come into contact. A surface nonplanarity of more than a few hundred angstroms results in unbonded areas and eventual leakage. Therefore, it is critical to thoroughly clean the bonding surfaces for a high yield process. Figure 9.26 shows a silicon substrate bonded to a glass capsule.

The glass capsule can be fabricated in several ways. It can be formed using molds of the right dimensions, or it can be batch fabricated using ultrasonic machining. Molding is time consuming, requires precision molds, and can provide variabilities from package to package that could compromise yield and bond quality. The glass capsule can be also formed by using ultrasonic machining of glass wafers. This technique is a batch process, provides precise control over the package dimensions, and can be used to form a variety of shapes and structures without the use of dedicated molds (although it does require a dedicated tool for ultrasonic machining).[37] The cavity of the capsule is milled by an ultrasonic machining tool. After milling, the wafer surface is polished to ensure planarity and smoothness. Finally, the polished wafer is diced apart to form the individual glass capsules. This method of forming the glass capsules has a number of advantages over the molding method. It is less expensive because the capsules are made in batch on a wafer level, and it provides a smoother and more planar surface because polishing is done on the wafer level. In addition, since the top and bottom surfaces of the capsule are flat, parallel and uniform mechanical pressure can be applied during bonding to enhance the bond quality. The top surface of the capsules can be metallized before dicing by standard wafer level metallization techniques to evenly distribute the electric field during bonding (with individually molded capsules, metallization is difficult). Finally, using a machined wafer allows for batch packaging of many systems at the wafer level. Figure 9.27 shows an SEM view of an ultrasonically machined glass capsule. Note that the features are well defined.

In addition to the glass capsule, one also needs to form the silicon substrate that supports the various electronics components, and to form multiple feedthroughs on the substrate to provide interconnection between sensors and the electronics. High-density on-chip multiple polysilicon feedthroughs are used for connecting external electrodes to the electronic circuitry inside the package. A high degree of planarity over feedthrough lines is a necessary condition for a reliable electrostatic bond. In order to achieve this planarity, closely spaced (2-μm separation) polysilicon lines, which are subsequently covered with a thick phosphosilicate glass (PSG), are used. The close spacing between the polysilicon lines allows the deposited dielectric to smoothly flow when subjected to a high-temperature ambient, and close the gaps, thus providing the planar surface. Figure 9.28 shows an SEM photograph of the cross section of feedthrough lines fabricated in this manner, showing the high degree of planarity. Although in the microstimulator application only two interconnect lines are required, in other applications that might require many more feedthrough lines this technique can be used to connect various other sensors to their associated interface circuitry inside the package. Note that the individual feedthrough lines can be easily connected in parallel to reduce the sheet resistance, which is desirable in some applications.

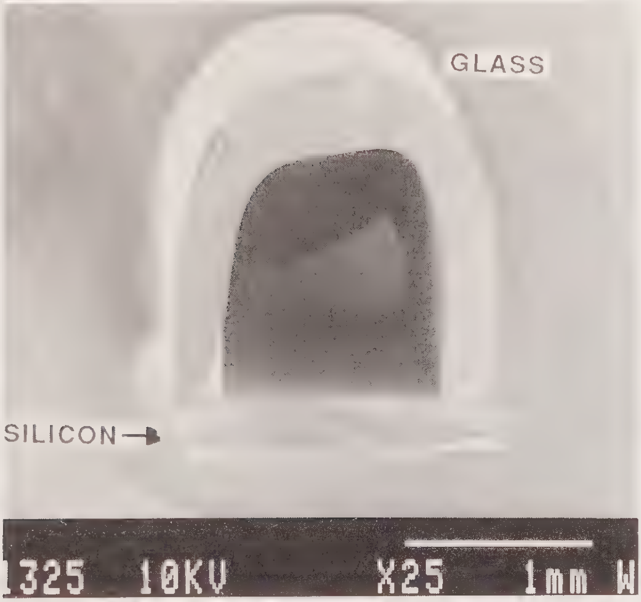

GLASS
SILICON
1325 10KV X25 1mm W

FIG. 9.26 Cross-sectional view of a silicon substrate bonded to a glass capsule.[37]

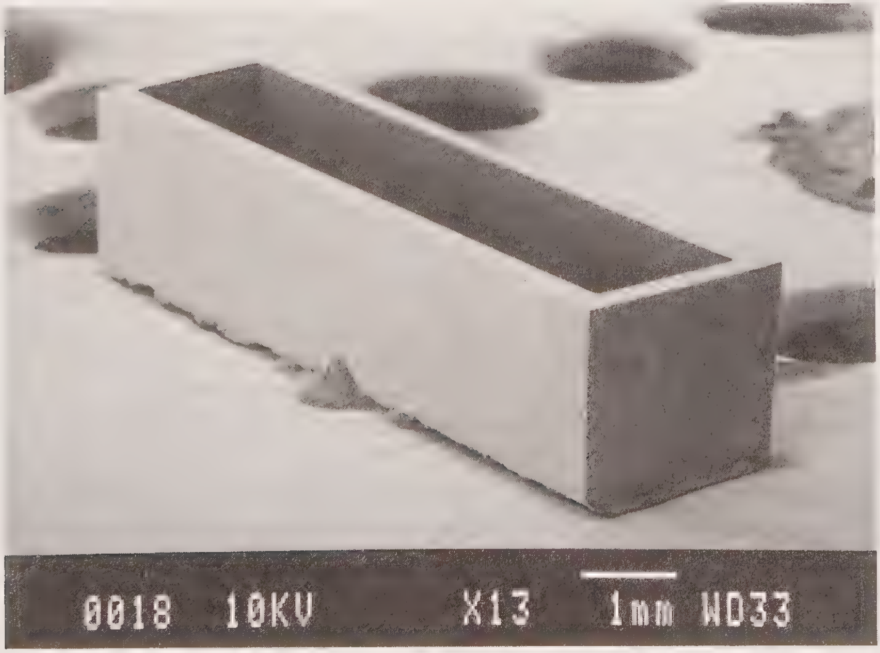

0018 10KV X13 1mm WD33

FIG. 9.27 SEM view of an ultrasonically machined glass capsule used for hermetic packaging.[37]

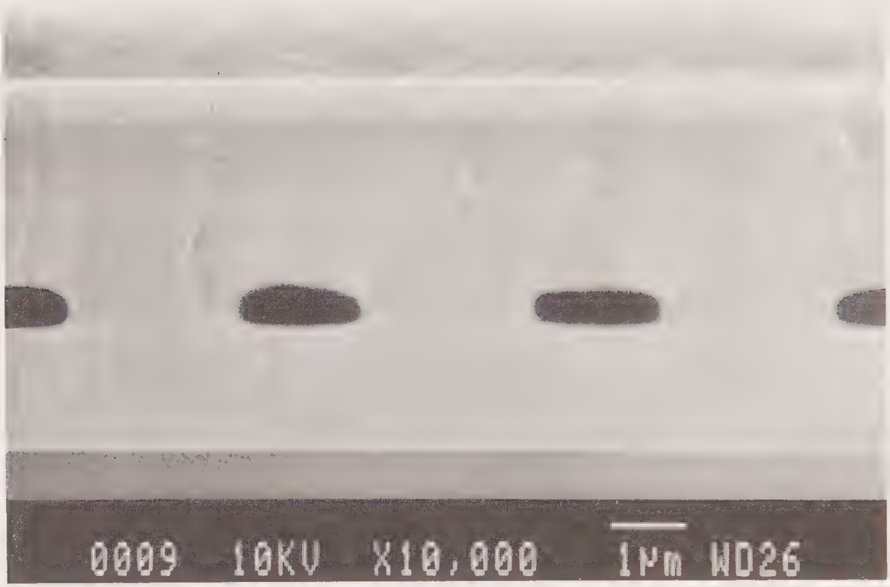
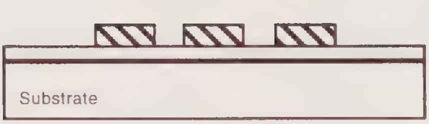
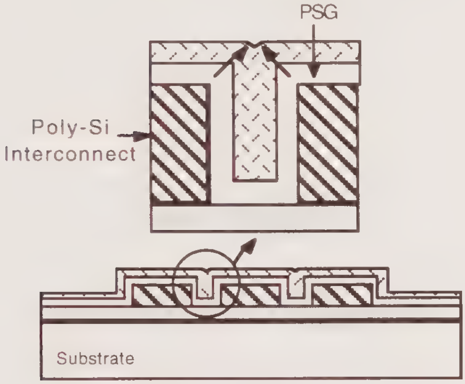


FIG. 9.28 SEM photograph of cross section of the grid structure feedthrough lines.[37]

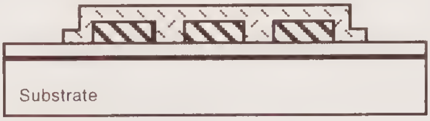
The fabrication process for the silicon substrate and the feedthroughs is shown in Fig. 9.29. Fabrication begins with a standard silicon substrate, over which ~1 μm of thermal silicon dioxide is grown, followed by the deposition of 1-μm-thick low-pressure chemical vapor deposited (LPCVD) polysilicon for feedthrough lines. The polysilicon is deposited at 570°C by pyrolyzing silane (SiH_4), and doped by diffusing phosphorus using $POCl_3$ liquid source at 950°C for 30 minutes. This yields a low sheet resistance needed for feedthrough lines (R_S≈5–10 Ω/square). The polysilicon is subsequently patterned using reactive ion etching (RIE) to produce the vertical sidewalls needed for planarization. This step is the most important masking step, and good photolithography is needed to achieve the closely spaced feedthrough lines. A combination of low temperature oxide (LTO) (deposited at 420°C) of about 2000 Å and phosphosilicate glass (PSG) (phosphorus concentration=8 wt.%) of about 1.2 μm is deposited on top of the polysilicon conductors. The LTO is required to improve the PSG adhesion to the first polysilicon. The PSG is then reflown at 1100°C in steam for 2 hours to planarize the top surface. Since PSG absorbs moisture rapidly, it is necessary to etch it everywhere except the areas over the feedthrough lines. The polysilicon conductor lines and the PSG are then insulated on top with a 3000 Å/1500 Å/3000 Å LPCVD silicon oxide, silicon nitride, silicon oxide sandwich to avoid their exposure to body fluids. Next the dielectrics (oxide/nitride/oxide, and the first oxide) are etched all the way down to the substrate. Since bonding the glass



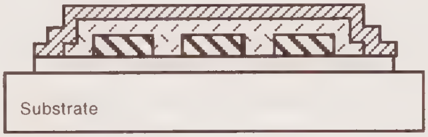
a) Grow the bottom insulating oxide. Deposit, dope, and pattern the first layer polysilicon.



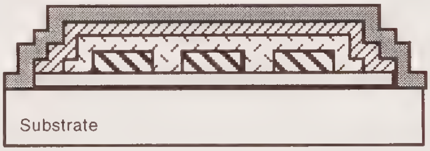
b) Deposit PSG on top of the first poly-Si layer and reflow at 1100 °C in steam to planarize.



c) Pattern the PSG and remove from everywhere except over the feedthrough lines and bond region.



d) Cap the PSG with LPCVD oxide-nitride-oxide film. Etch PSG & Dielectrics in the field regions.



e) Deposit and pattern the top, fine-grain polysilicon bonding layer.

FIG. 9.29 Fabrication process for the formation of the silicon substrate and on-chip feedthrough lines for the silicon-glass hermetic package.[37]

capsule to the oxide top layer requires a high temperature (~420°C) that could destroy hybrid components such as chip capacitors, a top polysilicon layer is deposited on the top dielectrics to reduce the bonding temperature (to about 320°C). The second polysilicon layer is 1 μm thick and is deposited similar to the feedthrough polysilicon lines, and is then doped using $POCl_3$ at 950°C for 30 minutes. This polysilicon layer wraps around the oxide/nitride/oxide layer and prevents its contact with moisture and therefore blocks water permeation through these layers inside the package. Finally, contacts are opened to the first polysilicon film, and a metal layer is deposited to achieve electrical contacts to the two polysilicon films.

Extensive accelerated soak tests have been performed on these packages. Several packages have been soaked in salt water at 95 and 85°C and the mean time to failure (MTTF) data has been used to calculate the expected MTTF at a temperature of 37°C. Based on these measurements we expect a lifetime of at least several decades for these packages, which is more than sufficient for many applications.

In addition to glass capsules, several investigators have also explored silicon capsules for creating a hermetically sealed cavity.[50,52] In this technique, a silicon capsule is bonded to another silicon substrate using a low-temperature gold-silicon eutectic bond. This technology can also be potentially used for packaging implantable systems. However, since in this technique the bonding temperature has to be raised to about 380°C, it may not be compatible with all hybrid components and limited to some applications. This technique is attractive because the silicon substrate can be more easily machined than the glass capsule and because the feedthroughs do not need to be perfectly planar since the gold-silicon eutectic flows at the bonding temperature and can cover over large steps. Before this technique can be used, however, long-term soak tests need to be conducted in salt water to verify the integrity of the bond and hermeticity of the package. It should also be mentioned that only a limited set of eutectic materials can be used in this approach because not all metals are biocompatible.

9.4.2 Nonhermetic Packaging Techniques

In addition to the use of packages like silicon and glass that provide a hermetic environment for the circuitry and electronic elements, it is also possible to package biomedical devices using deposited films. These films include inorganic thin films such as silicon nitride and silicon dioxide, or organic polymers such as silicone rubber, Parylene, and polyimides.[39] Silicon nitride is an excellent barrier against moisture penetration and has long been used by the IC industry to protect electronic elements from the ambient environment. It can be used to protect electronics and sensors from the harsh biological environments. It can be deposited at higher temperatures using a low-pressure CVD technique at lower temperatures using a plasma-enhanced CVD technique. The former has a much better quality, fewer pinholes, and excellent long-term stability, while the latter may require care in its deposition to make sure that pinholes are reduced. Long-term soak tests of LPCVD silicon nitride in saline solutions has shown excellent protection for a period of several years.[32]

A disadvantage of inorganic films such as silicon dioxide and silicon nitride is that they do not provide very good step coverage and cannot be used to encapsulate devices with hybrid components. In these devices, a number of polymeric films can be used. These include Parylene and various polyimides.[39] Parylene has long been used by the pacemaker industry and is still in widespread use and is an excellent insulator that is deposited at room temperature and is conformal. A number of polyimides and silicone rubbers have also been used to encapsulate biomedical devices. These polymers are very easy to use, do not require any dedicated equipment, and can be applied to a number of different surfaces. However, care has to be taken to ensure that they are applied carefully to devices they have to protect to make sure that good adhesion is achieved and delamination does not occur. It should be noted, however, neither organic nor inorganic films are fully hermetic and they will eventually leak when exposed to biological fluids for an extended period. Therefore, they should be used carefully in chronically implanted systems.

In summary, packaging of biomedical systems has long been a substantial challenge and continues to be a very important area of research and development. Packages that are small, biocompatible, low-cost, and simple to fabricate are needed in many applications, and new technologies are being developed to address some of these requirements.

9.5 CONCLUSION AND FUTURE TRENDS

As was presented above, solid-state technology has been crucial in the development of a number of sophisticated implantable recording and stimulating devices for interfacing with biological systems. Integrated circuit and micromachining technologies will continue to play an important role in the development of new recording and stimulation systems with enhanced capability and functionality.

In applications where information gathering and processing from the neural system is required, solid-state devices will be developed to further increase the number of recording sites in yet smaller volumes with more precise positioning and dimensions. The inclusion of on-chip circuitry not only improves the overall signal-to-noise ratio of the recorded signals, but also enables the user to program the system such that it can record only from the useful regions of the neural tissue under study. It is within the capabilities of today's technology to interface with the tissue through several hundred sites in a small volume of tissue over extended periods of time. In addition to recording information from the central nervous system, solid-state implantable sensors have been used for interfacing with the peripheral nervous system using sieve-type electrodes used in nerve regeneration applications. An important requirement for many of these applications is the transfer of data across the skin barrier. Silicon-based multi-lead ribbon cables have been developed and provide much needed interconnect systems, and have enabled the use of implantable microprobes in chronic situations. These cables are flexible, strong, and can support multiple interconnect leads in a small volume. The application of wireless telemetry techniques for transdermal data and power transfer is also becoming increasingly attractive and practical in miniature implantable systems. The marriage between miniature solid-state transducers and telemetry systems will simplify the packaging challenges facing implantable systems even further. It is anticipated that before the end of the century totally implantable recording/stimulation systems that are capable of interfacing with the tissue through several hundred sites in a chronic preparation will be developed.

In applications involving neural prostheses, solid-state technologies and transducers are even more important and critical. The successful development and realization of these systems requires the incorporation of a number of self-contained recording/stimulation units that are dispersed throughout the region of interest, and are electrically interconnected. Some units will record signals from the tissue and generate appropriate control signals for use by the stimulation units, which, in turn, deliver precise amounts of charge to the regions that need to be activated through appropriately designed stimulation circuits. These prosthetic systems will require the incorporation of sophisticated signal processing and control electronics in order to minimize data handling and control functions. Immediate application areas for these electrodes are in auditory and visual prostheses. Although stimulating electrodes and systems have already been developed for auditory prostheses, similar systems have not yet been successfully employed in visual prostheses. This is partly due to the increased complexity of these systems, which demand the development of sophisticated implantable units with large numbers of stimulating sites. The application of microelectronics and micromachining technologies similar to those discussed above is certain to remove some of the obstacles that researchers working on visual prostheses have faced for several decades.

The integration of electronics and transducer(s) for the development of implantable biomedical systems is critical for any future successes. Because of their difficult requirements and harsh operating environment, implantable systems pose many challenges on the fabrication technologies used for their implementation. These challenges are often simplified by adding on-chip circuitry to an otherwise passive transducer. The ability to amplify low-level biological signals on-chip, to multiplex several signals into a single output lead, thus reducing the total number of wires attached to the implant, and to buffer the output wires is very important to realizing reliable and stable implantable systems for a variety of applications. These circuit features often simplify the package requirements and reduce the long-term effects of prolonged exposure to harsh biological fluids. It is the conclusion of this paper that with today's technology, it is feasible to fabricate a variety of implantable sensors and actuators with on-chip circuitry and improve the overall performance of the system. While there are many factors that influence the partitioning of the system, the choice of the fabrication technology, the design of the particular circuit blocks, and the selection of appropriate materials for packaging, the fact remains that for most applications monolithic integration of transducers and circuitry is possible, desirable, and economically feasible. Examples of multichannel recording and stimulating microprobes for use in the central nervous system were given to illustrate the issues and demonstrate the existing capabilities.

Any implantable system for use in human beings has to be reliable, stable, and transparent to its user both aesthetically and functionally. All researchers working toward the development of these systems have to illustrate these features before these systems are put to actual use. There is no doubt that circuit incorporation and integration will continue to play an important role in the development of implantable biomedical systems. In fact, in some circles, as in groups involved in the development of closed-loop neural prostheses, it is believed that because of the enormous complexity of biological systems we will never have enough capability in terms of computational power, number of transducers, signal processing capability, small size, and speed, to truly match the performance of biological systems; no matter what we get we will need more. However, it has been demonstrated that implantable systems that are useful to people can and have been developed. Auditory implants that have helped bring some limited hearing to hearing impaired are an excellent example of this. These systems have been implanted in patients and are becoming more accepted as knowledge about their use increases. Many groups are working on the development of visual prostheses,[7] which is a much more formidable task. We need many more transducers, much faster operation, more sophisticated signal processing and data handling capability, lower power dissipation, and much better materials for packaging and interfacing with the biological tissue.

The continuing progress in the development of silicon-based microsystems will immensely benefit the implementation of practical electronic interfaces with biological tissues, and will eventually lead to the implementation of practical artificial prostheses that will help restore function where such function is lost. The same systems will also be used extensively by many researchers to enhance our knowledge of biological systems and their operation. Such knowledge is instrumental to successful development of man-made systems that provide similar performance.

ACKNOWLEDGMENTS

The research results and material presented in this chapter are based on research and development of solid-state integrated sensors carried out at the University of Michigan during the past fifteen years. A number of past and present students have participated in these research projects and their contributions are acknowledged. They include: Drs. Jin Ji (recording microprobes) and Steve Tanghe (stimulating microprobes), Dr. Tayfun Akin (sieve regeneration electrodes), Drs. Babak Ziaie and Mark Nardin (microstimulator and packaging), and Mr. Jeff Von Arx (packaging technology). The support of the staff at the Center for Integrated Sensors and Circuits of the University of Michigan is greatly appreciated. A major portion of the research and development reported here has been supervised and performed by Professor Kensall D. Wise, the director of the Center for Integrated Sensors and Circuits. The author is indebted to him for his inputs and contributions. The author also acknowledges the many contributions of Prof. David J. Anderson, Ms. J. Hetke, Prof. Robert M. Bradley, and Prof. S.L. BeMent of the Bioelectrical Sciences Laboratory at the University of Michigan in the development and application of these implantable systems.

The work reported here has been funded by a number of contracts and grants provided by the Neural Prosthesis Program, the National Institute for Neurological Disorders and Stroke, the National Institutes of Health. We are especially grateful for the continuing support and encouragement of Drs. Terry F. Hambrecht and William J. Heetderks of NIH.

REFERENCES

1. K.D. Wise, K. Najafi, "VLSI Sensors in Medicine," in *VLSI In Medicine*, Norman G. Einspruch and Robert D. Gold, eds., Academic Press, 1988
2. Special Issue of the *IEEE Engineering in Medicine and Biology,* Vol. 13, No. 3, June/July 1994
3. G. Lim, K. Minami, M. Sugihara, M. Uchiyama, and M. Esashi, “Active Catheter with Multi-Link Structure Based on Silicon Micromachining,”

Proceedings, IEEE Micro Electro Mechanical Systems Workshop, pp. 116-121, Amsterdam, The Netherlands, Feb. 1995

4. J. Ji, S.T. Cho, Y. Zhang, K. Najafi, and K.D. Wise, "An Ultraminiature CMOS Pressure Sensor for a Multiplexed Cardiovascular Catheter," *IEEE Trans. on Electron Devices*, Vol. 39, No. 10, pp. 2260-2267, October 1992
5. F.T. Hambrecht, J.B. Reswick (Eds.), in *Functional Electrical Stimulation: Applications in Neural Prosthesis,* New York: Marcel Dekker, 1977
6. *Neural Prosthesis: Fundamental Studies*, W. Agnew and D. McCreery, Eds., Englewood Cliffs, NJ: Prentice Hall, 1990
7. "Bioelectronic Vision, Special Report," *IEEE Spectrum*, May 1996
8. W. Gopel, "New Materials and Transducers for Chemical Sensors," *Sensors and Actuators,* B, 18-19, 1-21, 1994
9. H.S. Wong, and M.H. White, "A Self-Contained CMOS Integrated pH Sensor," *Technical Digest, Int. Electron Devices Meeting,* pp. 658-661, San Francisco, CA, December 1988
10. K. Najafi, K.D. Wise, "An Implantable Multielectrode Array with On-Chip Signal Processing," *IEEE J. Solid-State Circuits,* Vol. SC-21, No. 6, pp. 1035-1044, December 1986
11. T. Akin, B. Ziaie, K. Najafi, "RF Telemetry Powering and Control of Hermetically Sealed Integrated Sensors and Actuators," *Technical Digest, IEEE Solid-State Sensor and Actuator Workshop,* pp. 145-148, Hilton-Head, SC, June 1990
12. P.R. Gray, R.G. Meyer, "MOS Operational Amplifier Design- A Tutorial Overview," *IEEE J. Solid-State Ckts.*, Vol. SC-17, pp. 969-983, December 1982.
13. J. Kruger, "Simultaneous Recordings from many Cerebral Neurons: Techniques and Results," *Rev. Physiol. Biochem. Pharmacol.*, Vol. 98, pp. 177-233, 1983
14. G.E. Loeb, A.E. Walker, S. Uematsu, B.W. Konigsmark, "Histologic Reaction to Various Conductive and Dielectric Films Chronically Implanted in the Subdural Space," *J. Biomed. Mater. Res.*, Vol. 11, pp. 195-210, 1977
15. J. K. Niparko, R. A. Altschuler, X. Xue, J. A. Wiler, and D. J. Anderson, "Surgical Implantation and Biocompatibility of CNS Auditory Prostheses," *Annals of Otolaryngology*, vol. 98, pp. 965-970, 1989
16. J. K. Niparko, R. A. Altschuler, D. A. Evans, X. Xue, J. Farraye, and D. J. Anderson, "Auditory Brainstem Prosthesis: Biocompatibility of Stimulation," *Otolaryngology, Head Neck Surgery,* vol. 101, pp. 344-352, 1989
17. T. Tang, G. Ocvirk, and D.J. Harrison, "Iso-Thermal DNA Reactions and Assays in Microfabricated Capillary Electrophoresis Systems," *Digest, IEEE Int. Conf. on Solid-State Sensors and Actuators, Transducers '97*, pp. 523-526, Chicago, June 1997
18. K. Najafi, K.D. Wise, T. Mochizuki, "A High-Yield IC-Compatible Multichannel Recording Array," *IEEE Trans. on Electron Devices,* Vol. ED-32, No. 7, pp. 1206-1211, 1985

19. K. Najafi, J. Ji, K.D. Wise, "Scaling Limitations of Silicon Multichannel Recording Probes," *IEEE Trans. Biomed. Eng.*, Vol. BME-37, No. 1, p. 1, 1990
20. R.M. Finne, D.L. Kline, "A Water-Amine-Complexing Agent System for Etching Silicon," *J. Electrochem. Soc.*, Vol. 14, p. 965, 1967
21. A. Bohg, "Ethylene-Diamine-Pyrocatechol-Water Mixture Shows Etching Anomally in Boron-Doped Silicon," *J. Electrochem. Soc.*, Vol. 118, p. 401, 1971
22. A. Reisman, M. Bekenblit, S.A. Chan, F.B. Kaufman, D.C. Green, "The Controlled Etching of Silicon in Catalyzed Ethylenediamine-Pyrocatechol-Water Solutions," *J. Electrochem. Soc.*, Vol. 126, p. 1406, 1979
23. K. Najafi, J.F. Hetke, "Strength Characterization of Silicon Microprobes in Neurophysiological Tissues," *IEEE Trans. Biomedical Eng.*, 1990
24. A. Hoogerwerf, and K.D. Wise, "A Three-Dimensional Microelectrode Array for Chronic Neural Recording," *IEEE Trans. on Biomedical Engineering*, Vol. 41, No. 12, pp. 1136-1146, December 1994
25. Q. Bai, and K. Wise, "A High-Yield Process for Three-Dimensional Microelectrode Arrays," *Technical Digest, Solid-State Sensor and Actuator Workshop*, pp. 262-265, Hilton-Head Island, SC, June 1996
26. P.K. Campbell, K.E. Jones, R.J. Huber, K.W. Horch, and R.A. Normann, "A Silicon-Based, Three-Dimensional Neural Interface: Manufacturing Processes for an Intracortical Electrode Array," *IEEE Trans. on Biomedical Eng.*, Vol. BME-38, p. 758, 1991
27. J. Ji, and K.D. Wise, "An Implantable CMOS Circuit Interface for Multiplexed Microelectrode Recording Arrays," *IEEE J. of Solid-State Circuits*, **27**, No. 3, pp. 433-443, 1992
28. J. Ji, and K.D. Wise, "An Implantable CMOS Analog Signal Processor for Multiplexed Microelectrode Recording Arrays," *Technical Digest, IEEE Solid-State Sensor and Actuator Workshop*, pp.107-110, Hilton-Head, S.C., June 1990
29. D.W. Kewley, M.D. Hills, D.A. Berkholder, I.E. Opris, N.I. Maluf, C.W. Storment, G.T.A. Kovacs, "Plasma-Etched Neural Probes," *Technical Digest, Solid-State Sensor and Actuator Workshop*, pp. 266-269, Hilton-Head, S.C., June 1996
30. G. T. A. Kovacs, C.W. Storment, and J.M. Rosen, "Regeneration Microelectrode Array for Peripheral Nerve Recording and Stimulation," *IEEE Trans. on Biomedical Eng.*, Vol. BME-39, No. 9, pp. 893-902, 1992
31. T. Akin, K. Najafi, "A Micromachined Silicon Sieve Electrode for Nerve Regeneration Applications," *IEEE Trans. Biomedical Eng.*, Vol. BME-41, No. 4, pp. 305-313, April 1994
32. J.F. Hetke, J.L. Lund, K. Najafi, K.D. Wise, and D.J. Anderson, "Silicon Ribbon Cables For Chronically-Implantable Microelectrode Arrays," *IEEE Trans. Biomed. Eng.*, Vol. BME-41, No. 4, pp. 314-321, April 1994
33. T. Akin, K. Najafi, and R.M. Bradley, "An Implantable Multichannel Digital Neural Recording System for A Micromachined Sieve Electrode," *Proceedings, Int. Conf. on Solid-State Sensors and Actuators*, Volume 1, pp. 51-54, Stockholm, Sweden, June 1995;

Also to appear in: T. Akin, K. Najafi, and R.M. Bradley, "A Wireless Implantable Multichannel Digital Neural Recording System for a Micromachined Sieve Electrode," *IEEE J. Solid-State Circuits*, In Press, 1997

34. S.J. Tanghe, and K.D. Wise "A 16-Channel CMOS Neural Stimulating Array," *IEEE J. Solid-State Circuits,* 27, No. 12, pp. 1819-1825, 1992
35. L.S. Robblee, and T.L. Rose, "Electrochemical Guidelines for Selection of Protocols and Electrode Materials for Neural Stimulation," in *Neural Prosthesis: Fundamental Studies*, W. Agnew and D. McCreery, Eds., Englewood Cliffs, NJ: Prentice Hall, Chap. 2, 1990
36. B. Ziaie, Y. Gianchandani, K. Najafi, "A High-Current IrOx Thin-Film Neuromuscular Microstimulator," to be Published, *6th Int. Conf. on Solid-State Sensors and Actuators,* San Francisco, June 1991
37. B. Ziaie, J. Von Arx, M. Dokmeci, and K. Najafi, "Hermetic Glass-Silicon Micropackage with High-Density On-Chip Feedthroughs for Sensors and Actuators," *IEEE/ASME J. Micro Electro Mechanical Systems (J-MEMS),* Vol. 5, No. 3, September 1996
38. B. Ziaie, M. Nardin, J. Von Arx, and K. Najafi, "A Single-Channel Implantable Microstimulator for Functional Neuromuscular Stimulation (FNS*)," Proceedings, Int. Conf. on Solid-State Sensors and Actuators*, pp. 450-453, Yokohama, Japan, June 1993
Also to appear in: B. Ziaie, M.D. Nardin, A.R. Coghlan, and K. Najafi, "A Single-Channel Implantable Microstimulator for Functional Neuromuscular Stimulation," *IEEE Trans. on Biomedical Eng.*, In Press, To Appear in 1997
39. W. H. Ko, and T. M. Spears, "Packaging Materials and Techniques for Implantable Instruments," *Eng. Med. Biol.*, Vol. 2, pp. 24-38, March 1983.
40. L. Bowman, and J. Meindl, "The Packaging of Implantable Integrated Sensors," *IEEE Trans. on Biomed. Eng.*, Vol. BME-33, No. 2, pp. 248-255, Feb. 1986
41. G.E. Loeb, M.J. Bak, M. Salcman, E.M. Schmidt, "Parylene as a Chemically Stable, Reproducible Microelectrode Insulator," *IEEE Trans. Biomed. Eng.,* Vol. BME-24, No. 2, March 1977
42. P.E.K. Donaldson, and E. Sayer, "Silicon Rubber Adhesives as Encapsulants for Microelectronic Implants," *Med. Biol. Eng. and Comput.*, Vol. 15, 00.712-715, Nov. 1977
43. J.L. Lund, and K.D. Wise, "Chip-Level Encapsulation of Implantable CMOS Microelectrode Arrays," *Digest, IEEE Solid-State Sensor and Actuator Workshop,* pp. 29-32, Hilton-Head, SC, June 1994
44. M. F. Nichols, "The Challenges for Hermetic Encapsulation of Implanted Devices- A Review," *Critical Review in Biomedical Engineering*, 22(1):39, pp. 39-67, 1994
45. G.D. Wallis, D.I. Pomerantz, "Field-Assisted Glass-Metal Sealing," *J. Appl. Phys.,* Vol. 40, No. 10, pp. 3946-3949, September 1969
46. M. Esashi, A. Nakano, S. Shoji, H. Hebiguchi, "Low-Temperature Silicon-to-Silicon Anodic Bonding with Intermediate Low Melting Point Glass," *Sensors and Actuators,* A21-23 (1990), pp. 931-934

47. L.A. Field, R.S. Muller, "Fusing Silicon Wafers with Low Melting Temperature Glass," *Sensors and Actuators,* A21-23 (1990), pp. 935-938
48. B. Ziaie, J. Von Arx, M. Nardin, and K. Najafi, "A Hermetic Packaging Technology with Multiple Feedthroughs for Integrated Sensors and Actuators," *Int. Conf. on Solid-State Sensors and Actuators,* pp. 266-269, Yokohama, Japan, June 1993
49. J. Von Arx, B. Ziaie, M. Dokmeci, K. Najafi, "Hermeticity Testing of Glass-Silicon Packages with On-Chip Feedthroughs," *Proceedings, 8h Int. Conf. on Solid-State Sensors & Actuators,* pp. 244-247, Stockholm, June 1995
50. M.B. Cohn, Y. Liang, R.T. Howe, and A.P. Pisano, "Wafer-to-Wafer Transfer of Microstructures for Vacuum Packaging," *Technical Digest, Solid-State Sensor and Actuator Workshop*, pp. 32-35, Hilton Head, SC, June 1996
51. W. H. Ko, J. T. Suminto, and G. J. Yeh, "Bonding Techniques for Microsensors," in *Micromachining and Micro-packaging for Transducers,* Elsevier Science Publishers, B. V., 1985.
52. L.J. Spangler, and C.J. Kemp, "ISAAC- Integrated Silicon Automotive Accelerometer," *Digets, Int. Conf. on Solid-State Sensors and Actuators, Transducers '95*, Volume 1, pp. 585-588, Stockholm, Sweden, June 1995

CHAPTER 10

Thin-Film-Transistor-Addressed Liquid Crystal Displays

Fang-Chen Luo
Unipac Optoelectronics Corporation

CONTENTS

10.1 Introduction / 572

10.2 Device History / 572

10.3 Device Operation / 574

10.4 Semiconductor Materials / 579

10.5 TFT Array Fabrication Processes / 581
  10.5.1 TFT structures / 581
  10.5.2 Amorphous silicon TFT / 581
  10.5.3 Poly-Si TFT / 587
  10.5.4 Single crystal Si TFT / 590
  10.5.5 CdSe TFT / 592

10.6 TFT Array Inspection and Repair / 594

10.7 Color Filter Fabrication Processes / 595

10.8 LC Fabrication / 599

10.9 Optical Performance / 599

10.10 Display Applications and Status / 610

Conclusion / 612
References / 612

10.1 INTRODUCTION

During the past few years, active matrix liquid crystal displays (AMLCDs) or primarily thin-film-transistor-addressed liquid crystal displays (TFT-LCDs) have made significant progress in device performance and manufacturing technology. TFT-LCDs are well established in the portable PC application with world-wide capacity shortly to exceed two million units per month. The diagonal size of the displays has been enlarged steadily year by year up to 21 inches.[1] The resolutions of some displays have been increased to over 1000 lines per inch.[2] The manufacturing of these devices has matured with yields up to 80% reported. In this chapter, the development history, operation theory, fabrication processes, optical performance, and applications of these devices are discussed.

10.2 DEVICE HISTORY

An AMLCD incorporates an active matrix circuit in the LC cell to assist the electrical addressing of the cell. The AM circuit has an active device in every pixel element defined by the crossover of row and column bus lines. In a broad definition, the active device in an AM LC display has often been interpreted as any device with nonlinear characteristics. The nonlinear element can be either a two-terminal or a three-terminal device. The two-terminal devices for this application started in 1980 with the Ta_2O_5 MIMs[3] (metal-insulator-metal) and now include ring diodes,[4] back-to-back diodes,[5] p-i-n diodes,[6] and SiN_x MIMs.[7] The three-terminal devices are thin-film transistors (TFTs).

The two-terminal devices (or diodes) are generally considered an intermediate approach between a direct multiplexed LC cell and a three-terminal device because they have severe limitations in gray-scale performance and display size due to stringent requirements on device uniformity and threshold stability. Despite some potential advantages in simplicity and fabrication cost with two-terminal devices, three-terminal devices have become the selected technology and have received the predominant attention in the industry.

The invention of the TFT dates back to the 1930s when Lilienfeld was issued a U.S. Patent in 1933[8] and Heil a German patent a year later.[9] The first working TFT was realized by Weimer at RCA in 1962[10] using CdS as the semiconductor. There was a flurry of research activity[11-13] around the world on TFTs and their applications in the 1960s. However, interest faded away near the end of the decade when the Si metal-oxide-semiconductor (MOS) transistor established its superiority for applications in integrated circuits. The matrix circuit incorporating a field-effect transistor and a capacitor in every pixel element for the addressing of an LC display, which is still widely used today, was first reported by Lechner[14] in 1971. Fisher[15,16] et al. reported on the design of an LC color TV panel in 1972. The first attempt at constructing a TFT-LC panel was reported in 1973 by the Westinghouse group led by Brody,[17] which demonstrated the switching of one row of pixels in a 6×6-inch 20-line-per-inch (lpi) panel. The first fully operational TFT-LC alphanumeric panel[*] was achieved in 1974,[20] followed by the reporting of the video performance of a 6×6-inch 30 lpi TFT-LC panel in 1978,[21] all using CdSe. The performance of the 1978 panel is shown in Fig. 10.1. During the 1970s, the primary efforts on TFT-LC development were conducted at Westinghouse Research Laboratory using CdSe TFTs fabricated by the shadow mask technique[22] with display resolution limited to less than 50 lpi. The display resolution of CdSe TFT devices was increased to 50 lpi by a hybrid process combining shadow mask and photolithography[23,24] and to 80 lpi with 640×400 elements on an 8×5-inch active area by using an all-photolithographic process.[25]

Since the beginning of the 1980s, there has been an upsurge of interest in TFT-LC technology. This probably is due to the following reasons:

1. The availability of a-Si and poly-Si thin film as the semiconductor material for the TFT, which could utilize conventional semiconductor processing equipment for the fabrication of these devices.
2. The explosion in the development of personal laptop computers, which require a high- performance flat panel display.
3. The inadequacy and limitation of the performance of directly multiplexed LC displays.
4. The potential for the development of the pocket and personal TV market.

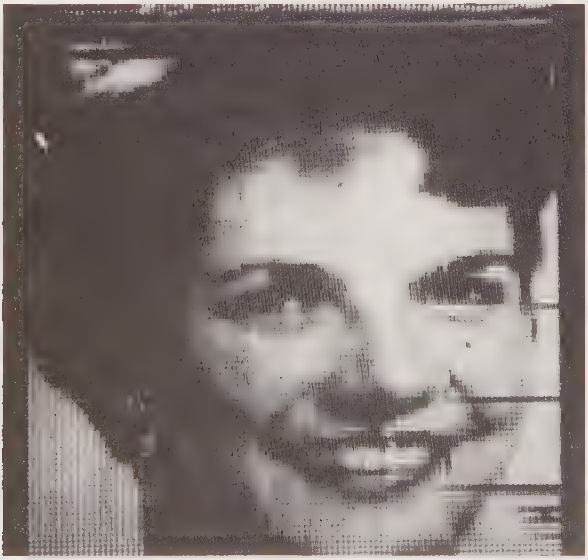


FIG. 10.1 Picture of a 1978 TFT-LC panel.

---

[*] A program on TFT-addressed electroluminescence (EL) panels was conducted in parallel with the TFT-LC effort at Westinghouse and resulted in the demonstration of 6×6-inch 20 lpi and 30 lpi TFT-EL panels in 1973[18] and 1975[19] respectively.

Currently, panel sizes up to 21 inches diagonal and panel resolutions over 1000 lpi have been reported. Portable PC displays are in high-volume production. An industry of TFT-LCDs has been established.

10.3 DEVICE OPERATION

As shown in Fig. 10.2, a color TFT-LC cell consists of a TFT matrix circuit glass plate and a color filter glass plate with the liquid crystal layer sandwiched in between. Each pixel element in the display cell is defined by the row and column bus lines in the TFT matrix circuit. The electrical circuit diagram of a pixel element is shown in Fig. 12.3. Each pixel element has one TFT and an LC capacitor, which is formed between the indium tin oxide (ITO) transparent output electrode on the TFT matrix circuit and the ITO backplane electrode on the color filter with the LC layer as the insulator. To operate the TFT-LC cell, a one-line-at-a-time addressing method is used. When a row (scan or gate line) is addressed, a positive pulse is applied to the line, turning on all TFTs along the row. The TFTs act as switches, transferring charges to LC capacitors from the respective columns (source, or data lines). When other rows are addressed, a negative voltage is applied to the gate line, turning off the TFTs along the line and holding the charges in the LC capacitors for one frame time until the line is addressed again. If the LC used in the cell is a twisted nematic (TN), it is desirable to use AC voltage to drive the LC element. The polarity of the data voltages switches in alternate frames. An example of the driving voltage waveforms of both the gate and source line is shown in Fig. 10.4.[27]

The TFT device works like a metal-oxide-semiconductor-field-effect transistor (MOSFET). The characteristics of a TFT generally can be represented by

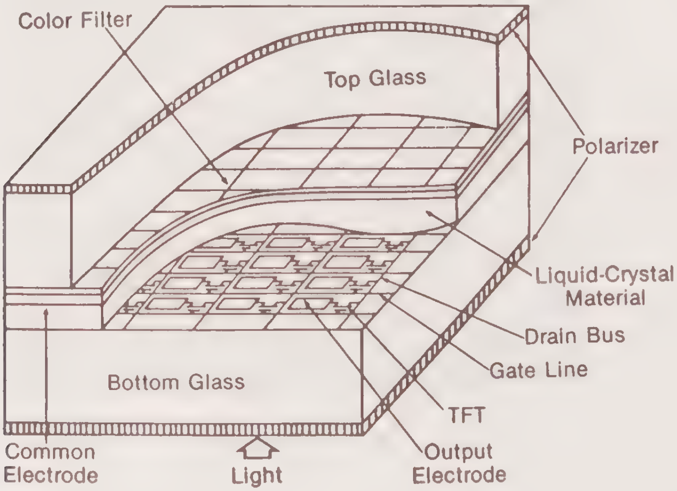

THE ANATOMY OF A TFT DISPLAY

FIG. 10.2 Cross-sectional view of a TFT-LC cell.[26]

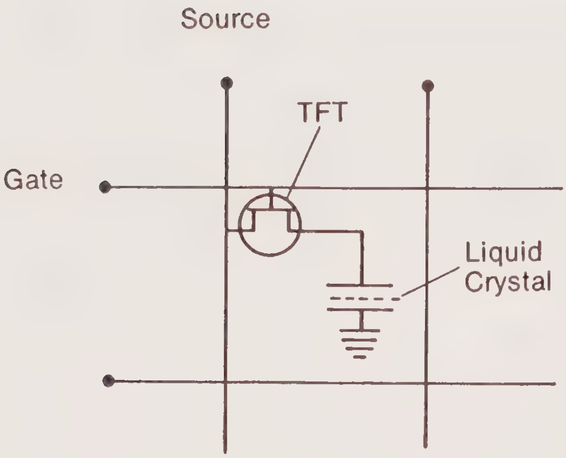


FIG. 10.3 Electrical circuit diagram of one pixel element.

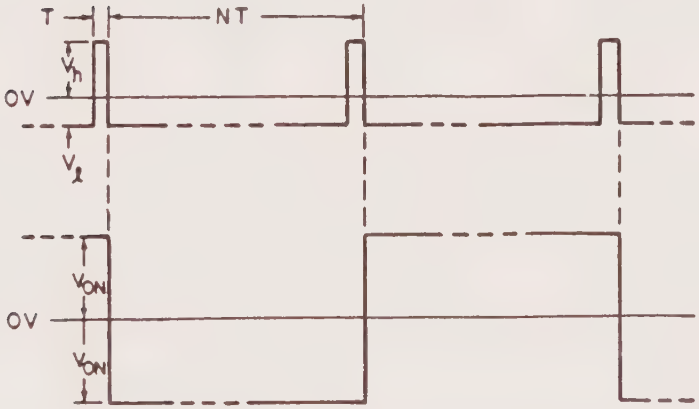


FIG. 10.4 Example of the driving voltage waveforms of both the gate and source lines.

$$
\begin{aligned}
&I_{SD} = (W\mu C_g/L)V_{SD}(V_G - V_T - V_{SD}/2)\\
&\text{when } V_{SD} \leq (V_G - V_T),\\
&\text{when } I_{SD} = (W\mu C_g/2L)(V_G - V_T)^2\\
&\text{when } V_{SD} > (V_G - V_T),
\end{aligned}
$$

where W is the width, L the length of the conducting channel of the TFT, μ the effective mobility of the semiconductor, C_g the gate capacitance per unit area, V_{SD} the source-drain (S-D) voltage, V_G the gate voltage, V_T the threshold voltage of the TFT, and I_{SD} the S-D current.

The requirement of the characteristics of the TFTs during the period is

$$I_{ON} > 2K_1 C_{LC} V_{ON} N/T,$$

where I_{ON} is the ON current of the TFT, K_1 is a constant, V_{ON} is the ON voltage of the LC element, C_{LC} is the capacitance of the LC element, N is the number of addressed lines in the displays, and T is the frame time. For a bilevel display such as an alphanumeric or graphic display, the requirement on the pixel voltage is not very tight since all voltage above the ON voltage of the LC element will be acceptable. K_1 can be a value between 1 and 10 to allow for any nonuniformity in the TFTs in the panel and the fact that the ON current of the TFT is not constant during the charging of the pixel.[27] For a 6.25×6.25-inch panel with 1024×1024 pixel elements, I_{ON} should be greater than 1.09×10^{-6}A if we pick K_1 to be 5.

For a gray-scale operation, the RC time constant is required to be at least one-third of the charging period to obtain less than 5% error. K_1 should be larger than 10. In this case, I_{ON} should be greater than 2.18×10^{-6}A. During the OFF periods, the TFTs are turned off. The OFF currents of the TFT should be low so the charges stored in the LC capacitors will not leak away to affect the appearance of the panel. For a bilevel panel, there are two separate conditions. For an OFF pixel, the voltage on the pixel should not increase to a level exceeding the threshold voltage of the LC material and turn the pixel partially ON. For an ON pixel, the voltage on the pixel should not drop below a voltage so the pixel appears partially ON or OFF. For a typical LC material this ΔV should be smaller than 1.5 V. The requirement on the OFF current of the TFT is

$$I_{OFF} < K_2\, \Delta V\, C_{LC}/T,$$

where K_2 is an engineering factor that allows for the variations of TFT characteristics in a TFT array and the temperature variations of the OFF currents of the TFTs. For a typical a-Si TFT, the OFF current can increase by a factor of 10 when the temperature is raised from 20°C to 70°C. Again using the 1024×1024 cell as an example, $\Delta V = 1.5$ V and $K_2 = 0.1$, I_{OFF} should be less than 3.28 pA.

For a gray-scale display, ΔV should be less than one gray level, or 0.08 V for a 16-level panel. I_{OFF} should be less than 0.175 pA.

Another path through which the charges stored in the LC capacitor can leak away is through the RC relaxation of the LC material itself.[28–30] For a 5% error of the root-mean-square voltage on the LC capacitor, the RC time constant of the LC material is required to be 10 times the frame time of addressing. For a display with a 16-ms frame time, the RC time constant should be longer than 160 ms. When considering the requirements of elevated-temperature performance of the cell, the room-temperature RC time constant should be over 1 s. In order to reduce the severe requirements on the TFT OFF currents and the LC RC time constant, a storage capacitor can be added to each pixel element.[14] The storage capacitor can be either a general group capacitor[31] or a capacitor formed between the ITO output electrode and the following scan line.[24] The trade-off in using a storage capacitor design is the increased fabrication complexity, reduced yield, and reduced aperture ratio. Between types, there is a difference in drive circuit options.

In a TFT-LC circuit, there are many parasitic capacitances, as shown in Fig.

10.5. C_{GD} is the capacitance in the TFT caused by the overlapping between the gate and drain electrodes. C_{GP} is the capacitance between the gate bus line and the ITO output electrode, which is much smaller than C_{GD}. During the addressing of a line, when the scan line voltage is dropped from V_{ON} to V_{OFF} as shown in Fig. 10.6, a negative voltage shift $\Delta V = \Delta V_G C_{GD}/(C_{LC}+C_{GD})$ results where $\Delta V_G = V_{ON} - V_{OFF}$. Because the geometry of the TFT array is not always uniform across the whole cell, and because C_{LC} is not a constant depending on the voltage applied to the LC element, this voltage shift cannot be compensated by a simple voltage shift of the backplane electrode. Excessive ΔV will result in the problems of flicker and degradation of gray-scale performance.

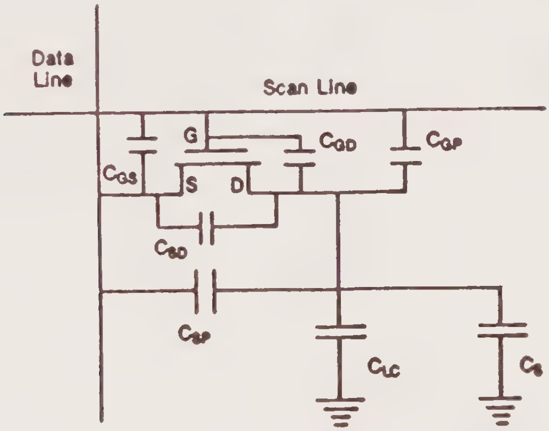


FIG. 10.5 Parasitic capacitances in a TFT-LC circuit.

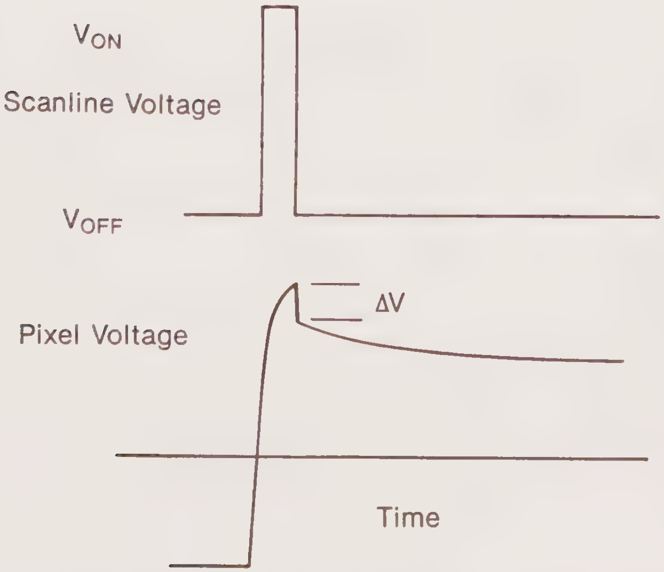


FIG. 10.6 Voltage shift in the pixel voltage caused by the scan line voltage.

Here again, a TFT array with a storage capacitor in each pixel has the advantage of reduced voltage shift from coupling. Because of the anisotropy of the LC material, C_{LC} is not constant, varying from a minimum of $C_{\perp}$ where no voltage is applied across the LC cell to a maximum of $C_{//}$ where the LC cell is fully turned ON. An example of a driving voltage compensation method[32] to achieve a symmetrical driving voltage across the LC cell is shown in Fig. 10.7. In this example, V_D is the data voltage applied to the data line, V_p is the voltage to the LC pixel electrode, V_{com} is the voltage of the common electrode, $dV_{p\perp}$ is the voltage shift when the LC is OFF, and $dV_{p//}$ is the voltage shift when the LC is ON. The solid line is the pixel voltage before the voltage shift and the dotted line is after the voltage shift. The data voltage $V_D = V_{sig\text{-}c} + V_{+amp}$ in one frame and $V_D = V_{sig\text{-}c} - V_{-amp}$ in the next frame. Typically, to obtain a symmetrical voltage across the pixel in all frames V_{+amp} is smaller than V_{-amp} and $V_{com} = V_{sig\text{-}c} - dV_{p\perp}$.

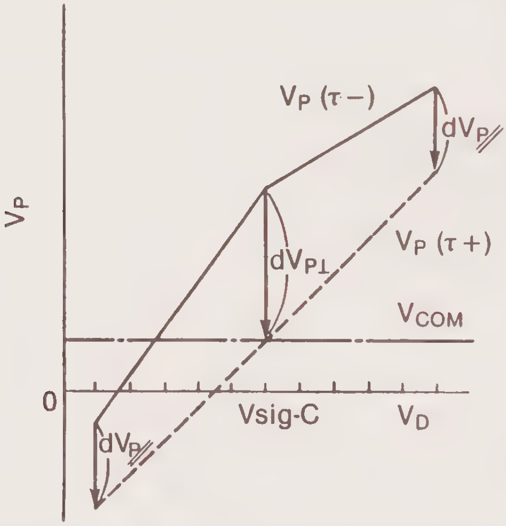


FIG. 10.7 Voltage compensation method to obtain symmetrical LC driving voltage.[32]

Another major source of gray-scale voltage distortion is through the capacitance coupling from the data voltages. The factor of the capacitor coupling is $(C_{SP}+C_{SD})/(C_{SP}+C_{SD}+C_{LC})$, where C_{SP} is the capacitance between the data line and the ITO output electrode and C_{SD} is the S-D capacitance of the TFT. For a 6-V data voltage, this voltage shift could be in the order of 0.25 V for a typical 164-lpi display design. Both neighboring data lines make contributions to this capacitance-coupling effect. One example of this effect is a top-to-bottom brightness variation in the gray-scale operation, since the top of the cell sees the reverse-polarity data voltage for a shorter time than the bottom part of the cell. To compensate for this voltage shift, two methods have been proposed.[33,34]

The first method[33] is the alternation of the data voltage polarity line by line and frame by frame. The second method[33] is to divide the addressing period of each line into two halves. In the first half, a normal data voltage V_i is applied. In the second half, the TFTs in the row are turned off and a data complement voltage ($V_m - V_i$) is applied to the data line where V_m is a constant voltage.

As mentioned earlier, through the capacitance coupling between the gate and drain electrodes of a TFT and the gate line and the ITO electrode, the driving voltage across the LC cell is shifted by an amount that can be up to 4 V in a typical 164-lpi TFT-LC cell. This LC voltage-level shift can cause a net dc voltage on the LC element. When the driving voltage across an LC cell has a dc component, the transmission output of the cell has an asymmetry between the frame that has a positive LC voltage and the next frame that has a negative LC voltage. If the refresh rate of the cell is low, this can result in a visible flicker as shown in Fig. 10.8. One way to solve the problem is to increase the refresh rate to greater than 100 Hz so the flicker will not be visible.[35] Another method is using the line inversion technique mentioned earlier.[33] An even more effective method is a combination of line inversion and column inversion, called pixel inversion.[36] The eye responds to the average modulation, which is reduced by these techniques. An example of an optical response using the pixel inversion technique is shown in Fig. 10.9.[36]

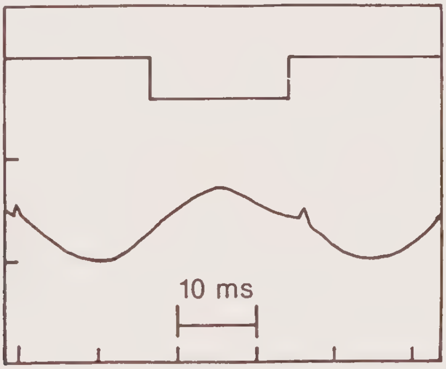


FIG. 10.8 Flicker in a cell.[32] The upper trace is the driving voltage, the lower trace the optical response of a pixel.

10.4 SEMICONDUCTOR MATERIALS

There have been four types of semiconductors utilized in TFT-LCDs: a-Si, poly-Si, single-crystal Si, and CdSe. CdSe has the longest history and has been applied to TFT-addressed displays since the 1970s.[15,37] It has the advantages of being a low-temperature processing material and having a relatively high mobility of up to 300 cm^2/V-s.[38] In the early days, CdSe devices were fabricated by using shadow masks[22] that had the limitations of low resolution (<50 lines per inch) and particle contamination. The largest reported CdSe TFT-LCD was fabricated

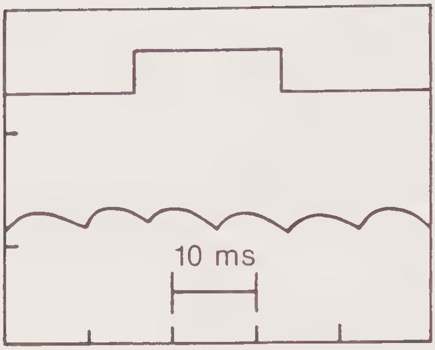


FIG. 10.9 Optical response of a flicker-free pixel.[32]

by an all-photolith process[39] on a 6×8-inch glass with a resolution of 160 lines per inch and a pixel count of 960×1280. CdSe has also been successfully used in display row and column drivers. A fully integrated 1.9-inch-diagonal cell with 288×480 pixels has been demonstrated.[40] The column driver included a shift register with seven TFTs per stage, working up to 7.5 MHz and a 4-bit switched capacitor D/A converter.

The development of a-Si TFTs has benefited significantly from the advancement of a-Si solar cells in the 1970s. Extensive knowledge on the deposition techniques and electrical characteristics was accumulated during those years. The material was first applied to TFTs in 1979[41] and has been extremely successful in TFT-LCD applications, primarily due to its high resistivity and adequate mobility for the pixel function. Amorphous silicon is noted for its good reproducibility and uniformity across a display. The manufacturing processes and equipment are well developed and mature. Process equipment to handle even larger substrates (currently up to 720 mm×600 mm) are becoming available every year. The largest-size display reported today is a 21-inch-diagonal 640×3×480 display.[1]

The a-Si TFT-LCD with the highest information content today is a 13.5-inch-diagonal monochrome display with 3072×2240 pixels.[42] Amorphous silicon TFTs have been utilized in projection displays also. Two examples[43,44] have 5.5-inch-diagonal 1024×1440 displays. Amorphous silicon TFTs have also been used in a fully integrated display. A 3.7-inch-diagonal 960×560 monochromatic projection display with fully integrated row and column drivers was successfully demonstrated.[45,46]

There are two types of poly-Si devices. One is a high-temperature (HT) poly-Si using quartz substrates and about 1000°C processing temperatures. The other is a low-temperature (LT) poly-Si processed on glass substrates under 600°C. Because of the size and cost limitation of quartz, HT poly-Si devices are focused on fully integrated small-area devices such as viewfinders and projection cells. An example is a fully integrated 1.8-inch-diagonal cell[47] with 1440×1024 pixels and a 22 μm×28 μm pixel dimension. An aperture ratio of 47.5% was achieved using the black-matrix-on-TFT array technology. The maximum clock frequency of the integrated shift register was 10 MHz at a 5-V applied voltage. Mobilities of up to

160 cm^2/V-s for n-channel and 110 cm^2/V-s for p-channel TFTs were achieved.[48] The HT poly-Si process is very mature. The device is manufactured in large quantities.

For the LT poly-Si devices, the Si films are deposited on glass substrates and recrystallized. Because of the glass substrate, the process has the potential to be applied to large areas.[49] Up to now, the largest reported direct-view display using LT poly-Si was a 9.5-inch 1140×1140 display.[50] An example of a fully integrated LT poly-Si device is a 0.7-inch-diagonal 420×220 device with a pixel dimension of 36 μm×46 μm.[51] The device has an aperture ratio of 27%. The operating frequency of the data driver was 1.3 MHz and the mobilities of the n-channel and p-channel TFTs were 38 cm^2/V-s and 30 cm^2/V-s respectively. Single-crystal Si devices have the best performance and the most mature manufacturing processes. They can perform the display multiplexing and integrated driving functions easily. Integrated crystalline silicon has the potential to integrate even higher level functions in the display chip. Of course, the size of the display is limited by the dimension of the Si wafer, and the cost per unit area is rather high.

Several single-crystal silicon processes for AMLCD applications have been investigated, including anisotropic etching of a single-crystal wafer,[52] polishing of a single-crystal wafer bonded to a quartz substrate,[53] and a transferred silicon process.[54,2] The transferred silicon process is the most developed, a 1280×1024 1000 pixel-per-inch AMLCD having been demonstrated.[2]

10.5 TFT ARRAY FABRICATION PROCESSES

10.5.1 TFT Structures

Two types of TFT structures have been commonly used in display applications. Fig. 10.10(a) is called an inverted staggered structure because the gate is at the bottom of the S-D and the gate electrodes are on the opposite side of the semiconductor. Amorphous silicon TFT arrays have used this structure extensively. Fig. 10.10(b) is called a coplanar structure because the gate and the S-D electrodes are on the same side of the semiconductor. This structure is popular with poly-Si TFT arrays. CdSe TFT arrays have used both structures.

10.5.2 Amorphous Silicon TFT

The fabrication processes for amorphous silicon TFTs can generally be classified into two types, A and B. The sequence of the fabrication process for a type-A device is shown in Fig. 10.11.[35] The starting substrate is usually a low-sodium-content hard glass such as Corning 7059 or 1737 glass. The gate material can be Ti, Ti-Mo, Cr, Ta, Al, Cu, or Mo-Ta, mostly deposited by dc magnetron sputtering. Ta and Mo-Ta gates have been widely used because the materials can be anodized, yielding a pinhole-free gate insulator.[55,56] Al with anodized Al_2O_3 has also been used as a gate.[57] Another method of reducing gate-line resistance, but retaining anodized Ta_2O_5, is to bury Al under Ta and Ta_2O_5.[58]

For the type-A process, after the gate electrode delineation, a sandwich of SiN_x,

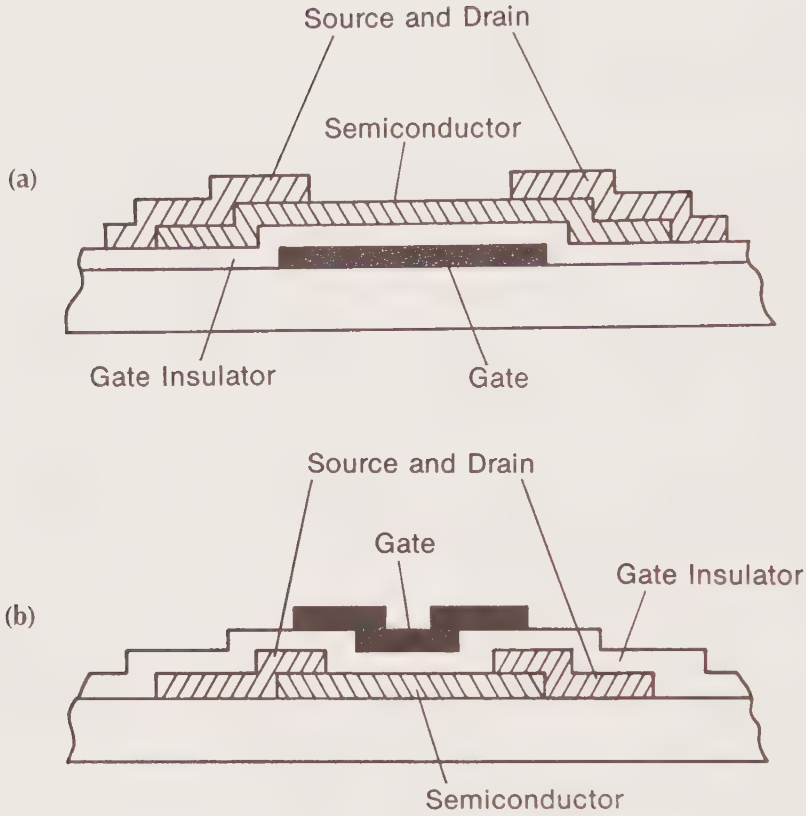


FIG. 10.10 (a) Inverted staggered TFT. (b) Coplanar TFT.

i-a-Si, and n+ a-Si is deposited in a one-vacuum pumpdown by the plasma-enhanced chemical vapor deposition process (PECVD). The layers of n+ a-Si and i-a-Si are etched into islands for individual TFTs. After the delineations of S-D and ITO electrodes, the n+ a-Si layer is etched to form the conducting channels for the TFTs using the S-D electrode as the mask. A timed n+ a-Si or a differential n+ a-Si etch stopping at the intrinsic a-Si layer is used to etch away the n+ a-Si layer and leave behind the intrinsic a-Si layer. Finally, a thin SiN_x layer is deposited on the array to passivate the devices.

For the type-B process, a typical process is shown in Fig. 10.12.[59] After the delineation of the gate, a sandwich of SiN_x, i-a-Si, and SiN_x is deposited in a one-vacuum pumpdown using PECVD. The top SiN_x layer is etched to expose the S-D regions. For a self-aligned TFT, the exposure is from the back side of the glass using the gate as a mask. Following the SiN_x etching, a thin layer of n+ a-Si is deposited in a PECVD machine and the n+/i-a-Si is patterned. After the deposition and patterning of the ITO pixel electrode, the S-D electrodes are delineated, connecting the TFT and the pixel electrode. The typical S-D materials are Cr, Mo, Cr-Al, Mo-Al, and Ti. Finally, a thin SiN_x layer is deposited by PECVD and etched to form the passivation. The self-aligned TFT still has some overlapping

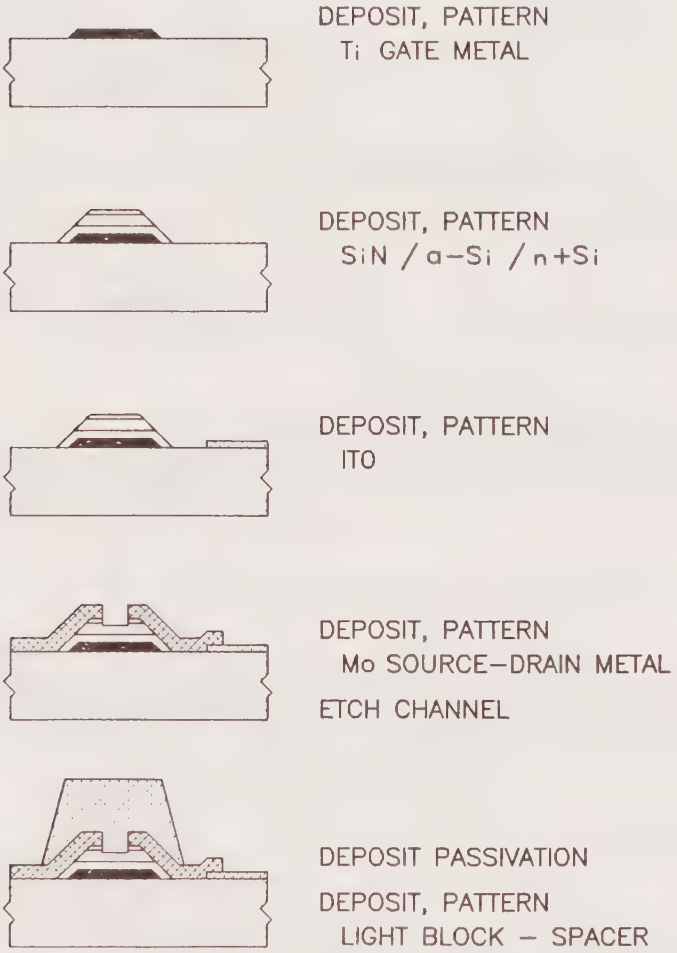

DEPOSIT, PATTERN
Ti GATE METAL
DEPOSIT, PATTERN
SiN / a–Si / n+Si
DEPOSIT, PATTERN
ITO
DEPOSIT, PATTERN
Mo SOURCE–DRAIN METAL
ETCH CHANNEL
DEPOSIT PASSIVATION
DEPOSIT, PATTERN
LIGHT BLOCK – SPACER

FIG. 10.11 Fabrication process of a type-A a-Si TFT array.[35]

between the gate and electrode and the S-D electrodes over the top insulator, which can cause a detrimental level shift. To further reduce this effect, a fully self-aligned gate process was developed.[60] The process steps are shown in Fig. 10.13.[60] Similar to the self-aligned TFT, the etch-stop SiN_x is exposed from the back side of the glass using the gate as the mask. After etching the SiN_x, the S-D regions are formed by ion doping to the 50-nm-thick i-a-Si film. Ten-percent PH_3 gas diluted with H_2 is discharged, and ions such as Ph_x+, H_y+ are accelerated and implanted into the substrate without mass separation. A layer of Mo is deposited to form Mo silicide. After patterning the a-Si island, Al-Mo electrodes are delineated, contacting the Mo silicide electrodes without overlapping the top insulator. The fully self-aligned TFT can reduce the gate-source-drain overlapping dimension from about 4 μm for a regular TFT to 0.5 μm and can make TFTs with shorter channel lengths feasible; therefore it can significantly improve the driving capability of the TFT array.

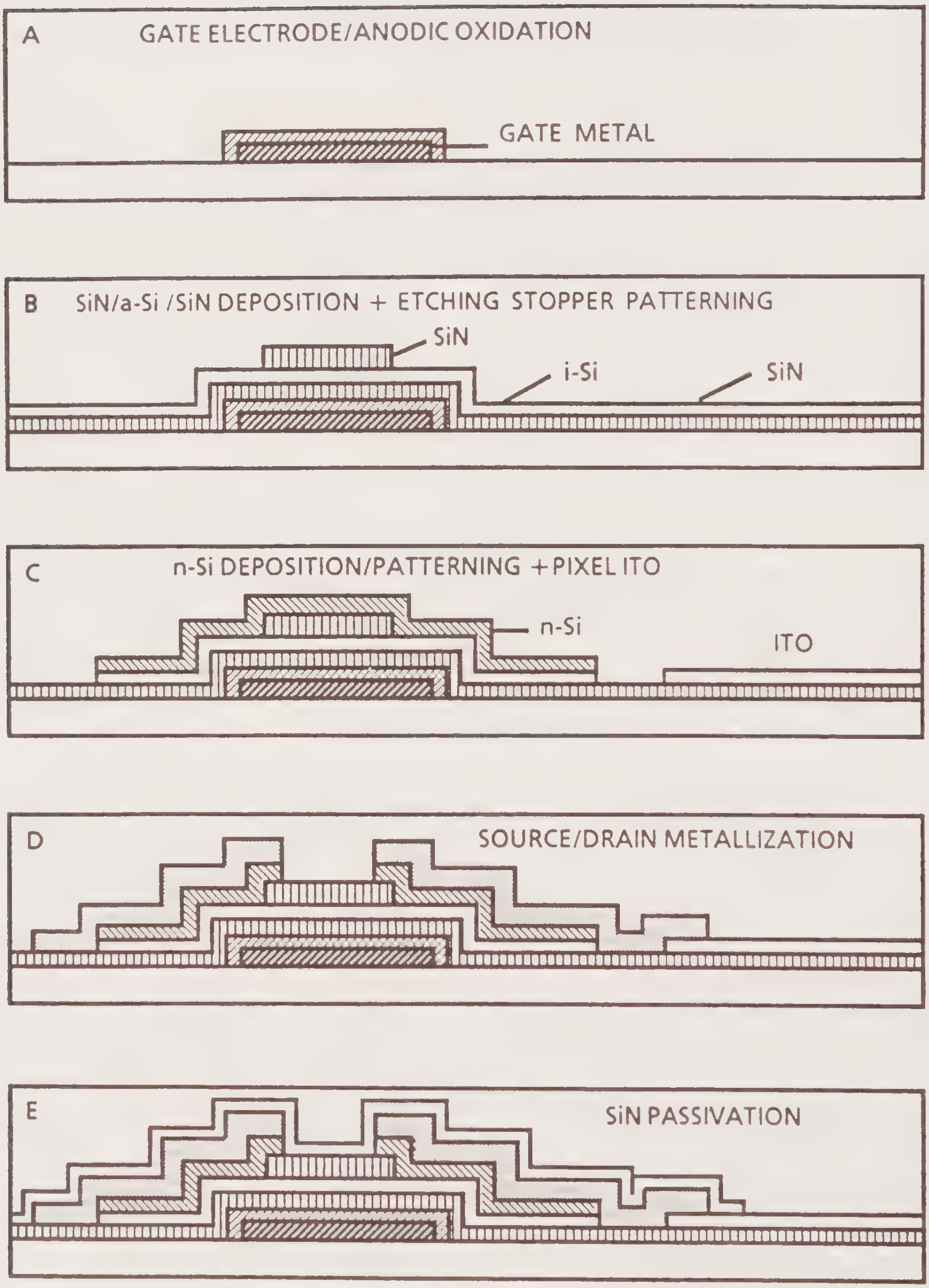
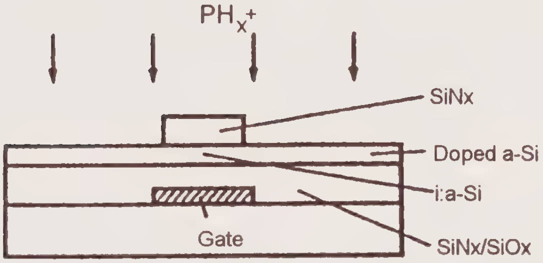
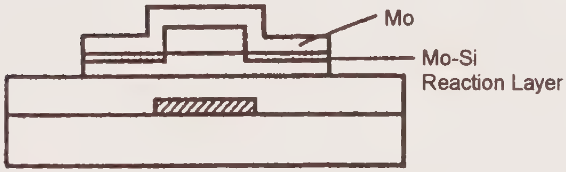

A
GATE ELECTRODE/ANODIC OXIDATION
GATE METAL
B
SiN/a-Si /SiN DEPOSITION + ETCHING STOPPER PATTERNING
SiN
i-Si
SiN
C
n-Si DEPOSITION/PATTERNING + PIXEL ITO
n-Si
ITO
D
SOURCE/DRAIN METALLIZATION
E
SiN PASSIVATION

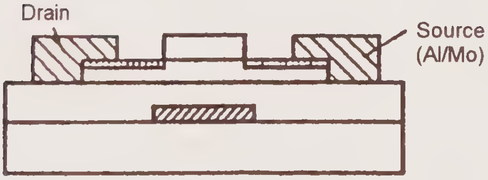
FIG. 10.12 Fabrication process of a type-B a-Si TFT.[59]



(a) DOPING



(b) SILICIDE FORMATION



(c) S/D PATTERNING

FIG. 10.13 Fabrication process of a fully self-aligned TFT.[60]

Comparing type-A and type-B processes, the type-B process requires five mask levels and needs careful cleaning of the interface before the deposition of the n+ a-Si layer to ensure obtaining good ohmic contacts for the TFTs. The type-A process only requires four mask levels; however, because of the lack of an efficient n+ a-Si etch stop, the thickness of the i-a-Si layer cannot be reduced to a few hundred angstroms to minimize the photoconductivity effect.[61] Also, in the type-A device, the passivation SiN_x layer is not deposited in the same pumpdown as the i-a-Si film, so the OFF current and the photocurrent of the TFT in general are higher than for type-B devices.[62] The I-V characteristics of a typical a-Si TFT are shown in Fig. 10.14.[62] An ON/OFF current ratio greater than 10^6 and mobility between 0.3 and 1 cm^2/V-s have been reported.

There are threshold voltage shifts in the TFT when dc biases are applied to the device. In general, as shown in Fig. 10.15, the threshold voltage is shifted to positive under a positive bias and to negative under a negative bias.[63] This is

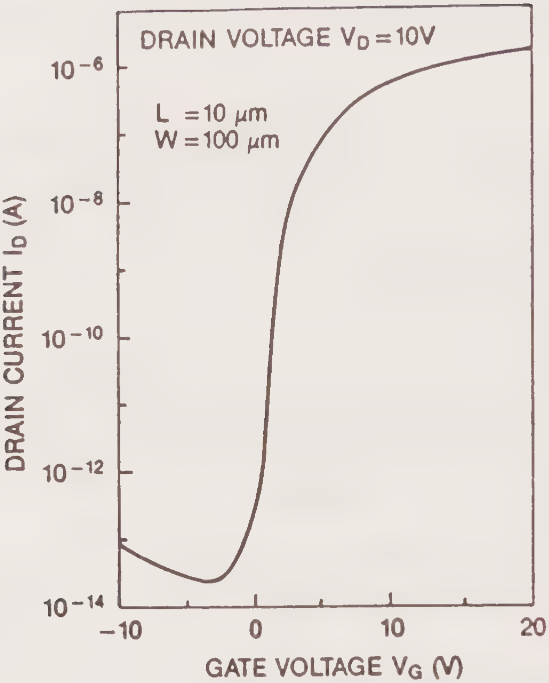


FIG. 10.14 Characteristics of a typical a-Si TFT.[62]

caused by charge injection into the gate insulator layer. The variable-range hopping model for the conduction in SiN_x was proposed by Powell.[64] Suzuki et al. showed that the voltage shift corresponds to the summation of positive and negative biased periods.[65] However, in a dynamic TFT-LC operation, due to the compensating forces of the alternating positive and negative voltages applied to

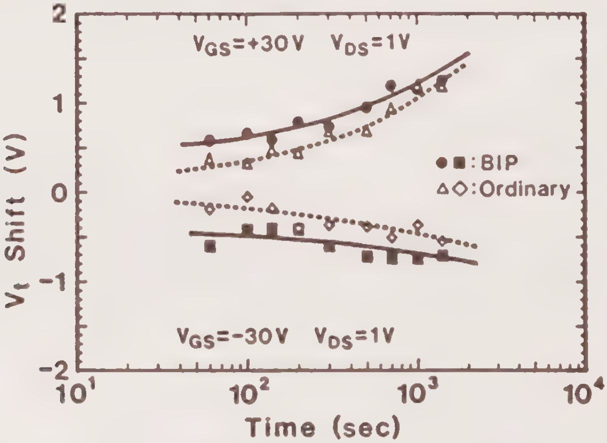


FIG. 10.15 Threshold voltage shift of an a-Si TFT under dc biases.[63]

the TFT, stable operations can be obtained.[66] Figure 10.16 shows the changes of the threshold voltage of an a-Si TFT vs. time in various TFT-LC operations at elevated temperatures.[66] An operation of more than 1.93×10^8 s, or 6.1 years for an XGA display (1024×768 pixels), at 50°C is obtained if a limit of 2 V for the TFT threshold voltage shift is assumed.

10.5.3 Poly-Si TFT

The substrates used in the HT poly-Si process are quartz. An example of a process is shown in Fig. 10.17.[67] At first, a poly-Si film is deposited on the substrate by LPCVD and is patterned and then oxidized at 1000°C, forming 70-nm-thick SiO_2 as the gate oxide. Then boron ions are implanted into the TFT channel regions for TFT threshold voltage control and phosphorus ions are implanted into the electrode regions for pixel storage capacitors [Fig. 10.17(a)]. The gate electrodes are formed by an Mo-polycide process.[68] N-type and P-type impurities are successively implanted into the N-, N+, and P+ regions [Fig. 10.17(b)] forming TFT contacts by the self-aligned process for the n-channel and p-channel and lightly-doped-drain (LDD) n-channel TFT.

Subsequently, an 800-nm-thick SiO_2 layer is deposited on the substrate and annealed at 1000°C. After hydrogenation, contact via holes are reactive ion etched (RIE) in the SiO_2 layer and Al data lines are deposited and delineated as shown in Fig. 10.17( ). Then the second 800-nm SiO_2 layer is deposited and the second contact via holes are RIE etched. Finally, transparent ITO electrodes are deposited and patterned as the pixel electrodes. In this example, the pixel electrodes are insulated from and overlap the data and scanning lines. The data and scanning lines serve as the black matrix and produce an aperture ratio of 35% in a 14μm ×42-μm pixel pitch design.

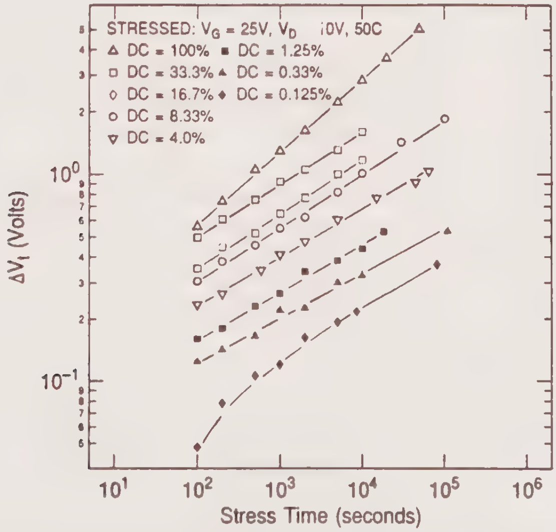


FIG. 10.16 Threshold voltage shifts vs. stress time for various duty cycles.[66]

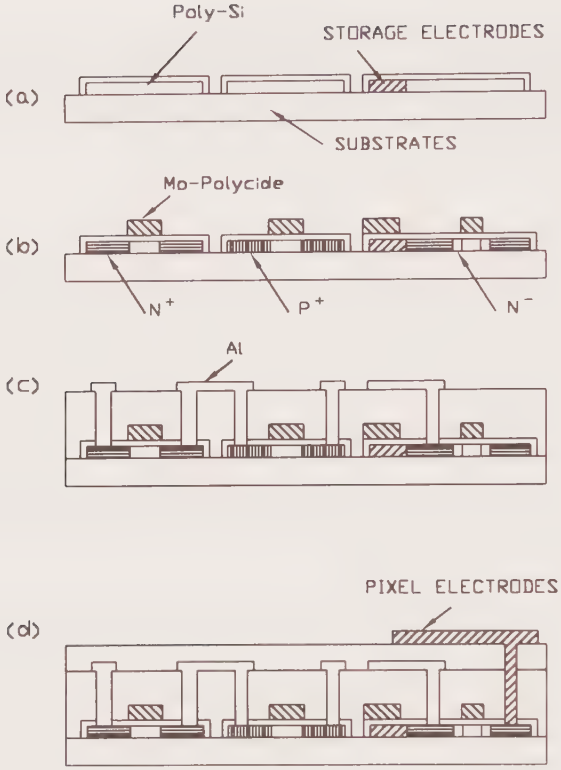


FIG. 10.17 An HT poly-Si TFT process.[67]

TFTs with the LDD structure have low leakage current[69] without resorting to a dual-gate structure. Figure 10.18 shows the characteristics of an LDD TFT with a width-to-length ratio of 3/5 μm.[67] Figure 10.19 shows the characteristics of n-channel and p-channel TFTs for the drivers.[67] Mobilities of 101 and 70 cm^2/V-s for the n-channel and the p-channel TFT, respectively, are obtained. Shift registers for the data drivers operate at a frequency of 7.84 MHz at 12 V V_{DD}.

In LT poly-Si processes, the substrates are borosilicate glasses and the silicon films require recrystallization. Figure 10.20 shows the sequence of a typical process.[51] The a-Si film is deposited by LPCVD on a glass substrate (NA35) with an SiO_2 barrier layer and is recrystallized by XeCl excimer laser in two steps. After patterning the TFT channel regions, a 120-nm-thick SiO_2 layer is deposited by ECR CVD in a vacuum. The electrodes of the pixel storage capacitor are formed by a non-mass-separated ion implanter using photosensitive polyimide as the mask and phosphorus ions as the dopant. After the Ta gate delineation, phosphorus ions and boron ions are implanted into the silicon layer, forming source and drain electrodes for the n-channel and p-channel TFTs, respectively, using the self-aligned process. Impurities are activated at 300°C for 2 hours in an N_2 atmosphere. After depositing a thick SiO_2 layer, via holes are opened at the source-drain regions. Al electrodes are delineated as data lines and ITO electrodes as pixel electrodes.

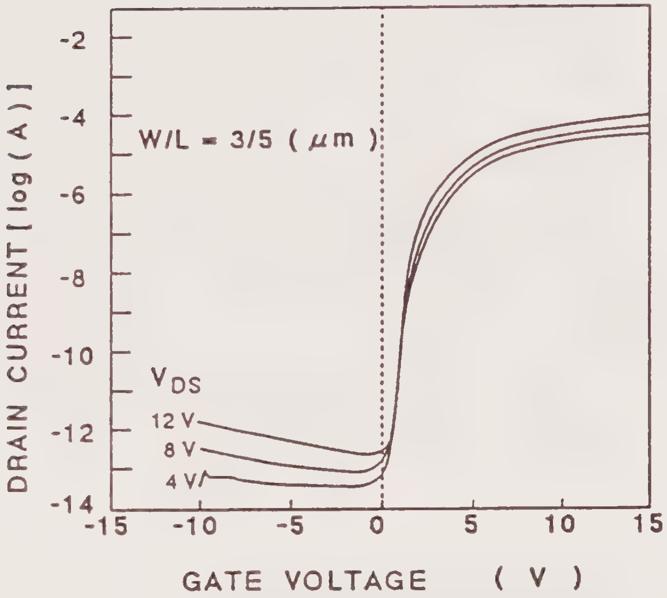


FIG. 10.18 Characteristics of an HT poly-Si pixel TFT.[67]

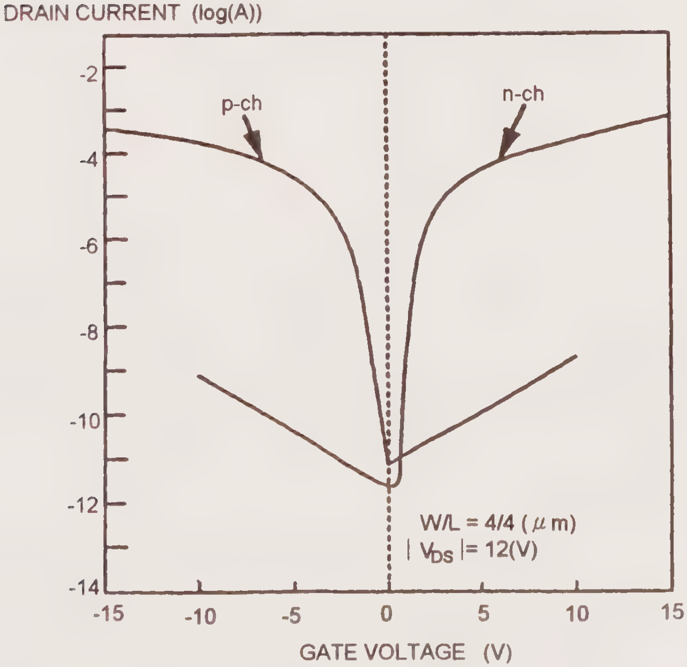


FIG. 10.19 Characteristics of HT poly-Si driver TFTs.[67]

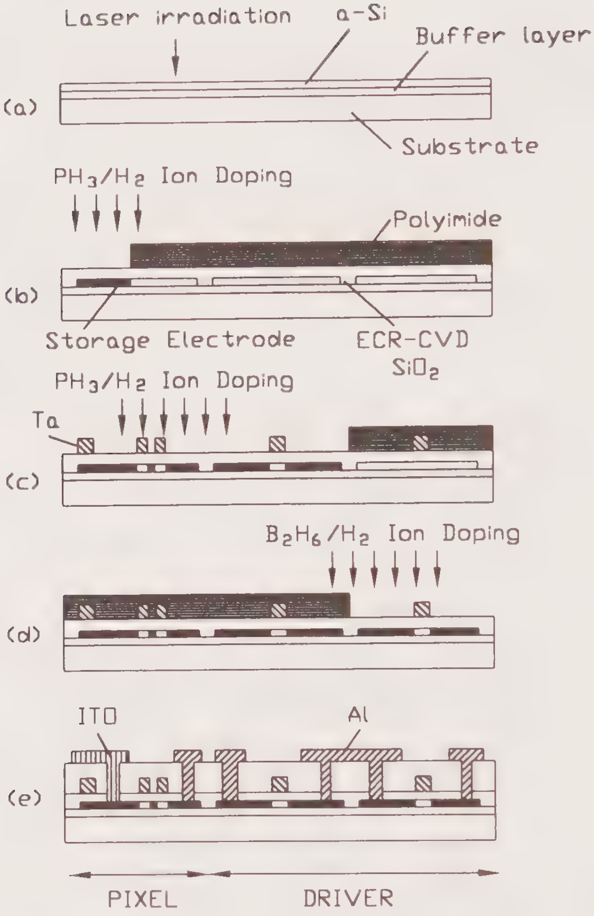

Laser irradiation
a-Si
Buffer layer
(a)
Substrate
PH3/H2 Ion Doping
Polyimide
(b)
Storage Electrode
ECR-CVD SiO2
PH3/H2 Ion Doping
Ta
(c)
B2H6/H2 Ion Doping
(d)
ITO
Al
(e)
PIXEL
DRIVER

FIG. 10.20 An LT poly-Si TFT process.[51]

The TFT characteristics of the n-channel and p-channel TFTs prepared by this process are shown in Fig. 10.21.[51] Dual-gate TFTs are used with W/L dimensions of 3 μm/8 μm+ 8 μm as the pixel TFT. For a pixel pitch of 34 μm×46 μm, an aperture ratio of 27% is achieved. The mobilities for the n-channel and p-channel TFT for the integrated drivers are 38 cm^2/V-s and 30 cm^2/V-s respectively. The operating frequency of the data driver is 1.3 MHz.

10.5.4 Single-Crystal Si TFT

The transferred silicon process is the most developed single-crystal Si process, having been used to demonstrate a 1280×1024^2 1000 pixel-per-inch TFT-AMLCD. Figure 10.22 shows the process sequence.[54] First, a 0.3-μm-thick x-Si layer is formed on a thermally oxidized Si wafer by an isolated silicon epitaxy process in which a vapor-deposited poly-Si thin film encapsulated by an oxide cap layer is recrystallized by a scanning hot filament and is seeded from the original x-Si wafer through a hole in the buried oxide. After removing the oxide cap and patterning, the Si islands are thermally oxidized to form a 70-nm-thick gate oxide. Boron and phosphorus ions are implanted into the islands for threshold voltage control of n- and p-channel MOS transistors. A layer of LPCVD polysilicon is then deposited at 560°C to a thickness of 0.5 μm and patterned to form the transistor gates. Phosphorus and boron implants are used to dope the source/drain regions, and an APCVD boron-phosphorus-silicon-glass (BPSG) is deposited over the surface to form the interlayer dielectric. The wafers are then heated to 850°C for 20 minutes to flow the BPSG and activate the implants. Contact paths are opened in the glass and reflowed, and aluminum metal is de posited and defined to form the interconnect pattern. A silicon nitride layer is deposited over the surface of the wafers prior to 450°C contact sintering. After array processing, the wafers are neutron irradiated in order to minimize effects associated with the high carrier lifetime in the x-Si film.

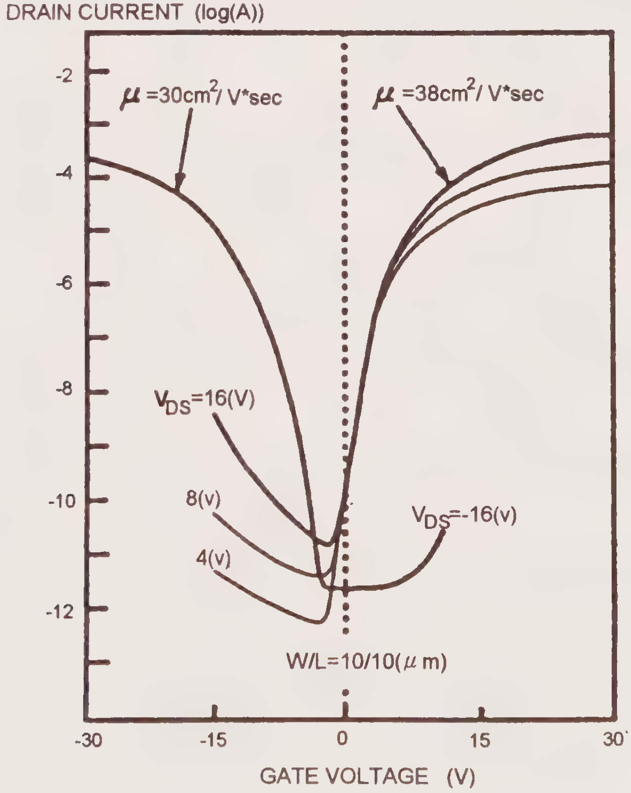


FIG. 10.21 TFT characteristics of n-channel and p-channel LT poly-Si TFTs.[51]

After circuit processing, the 2.5×2.5-inch integrated active matrix display circuit is removed from the silicon substrate and permanently bonded to glass using a thick layer of adhesive. Both Corning 7059 and Hoya NA-40 glass have been used. As a consequence of the transfer process, the original top surface of the

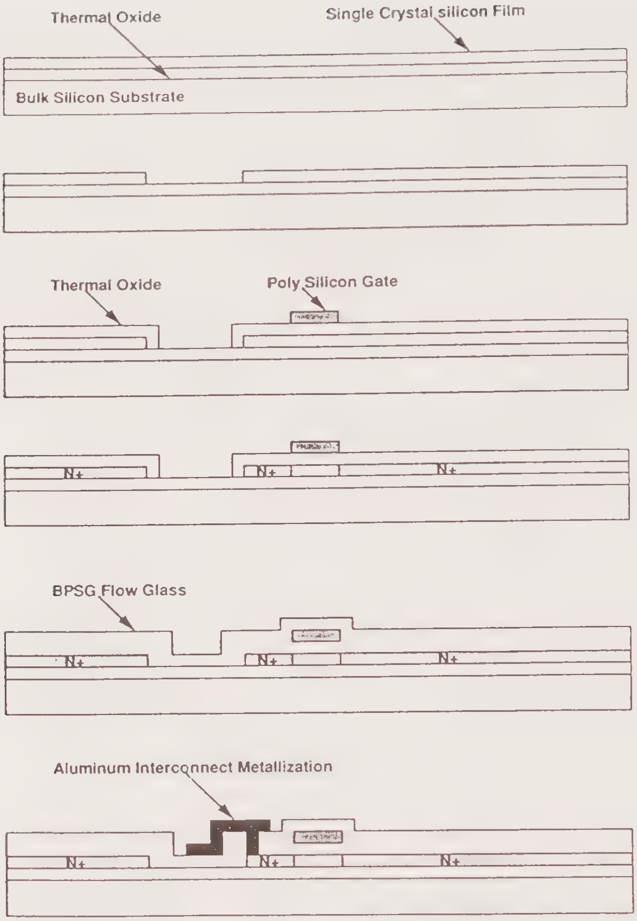

Thermal Oxide
Single Crystal silicon Film
Bulk Silicon Substrate
Thermal Oxide
Poly Silicon Gate
N+
N+
N+
BPSG Flow Glass
N+
N+
N+
Aluminum Interconnect Metallization
N+
N+
N+

FIG. 10.22 Transferred silicon TFT fabrication process.[54]

circuit is placed adjacent to the glass and the back side of the circuit is exposed. Photolithographic processing is used post-transfer to expose the input contacts. The x-Si circuit is then put through a final test and the laser repair procedure normally done to the equivalent poly-Si circuits. Excellent characteristics are obtained. For a 6-μm-L×10-μm-W device, the minimum leakage current is 0.06 pA/μm. The extrapolated threshold voltage for the n-channel device is 2.0 V and the low field effect mobility is 467 cm^2/V-s. The p-channel device has a threshold voltage and mobility of minus minus 1.5 V and 195 cm^2/V-s, respectively.

10.5.5 CdSe TFT

A representative CdSe TFT process is shown in Fig. 10.23.[70] The gate is sputter-deposited, 130-nm-thick, wet-etched Cr. The gate insulator is 450-nm-thick SiO_2, deposited by APCVD. After exposing the Cr contact pads by plasma etching, the substrate is sputter-etched and coated with a thin CdSe layer by thermal evapo-

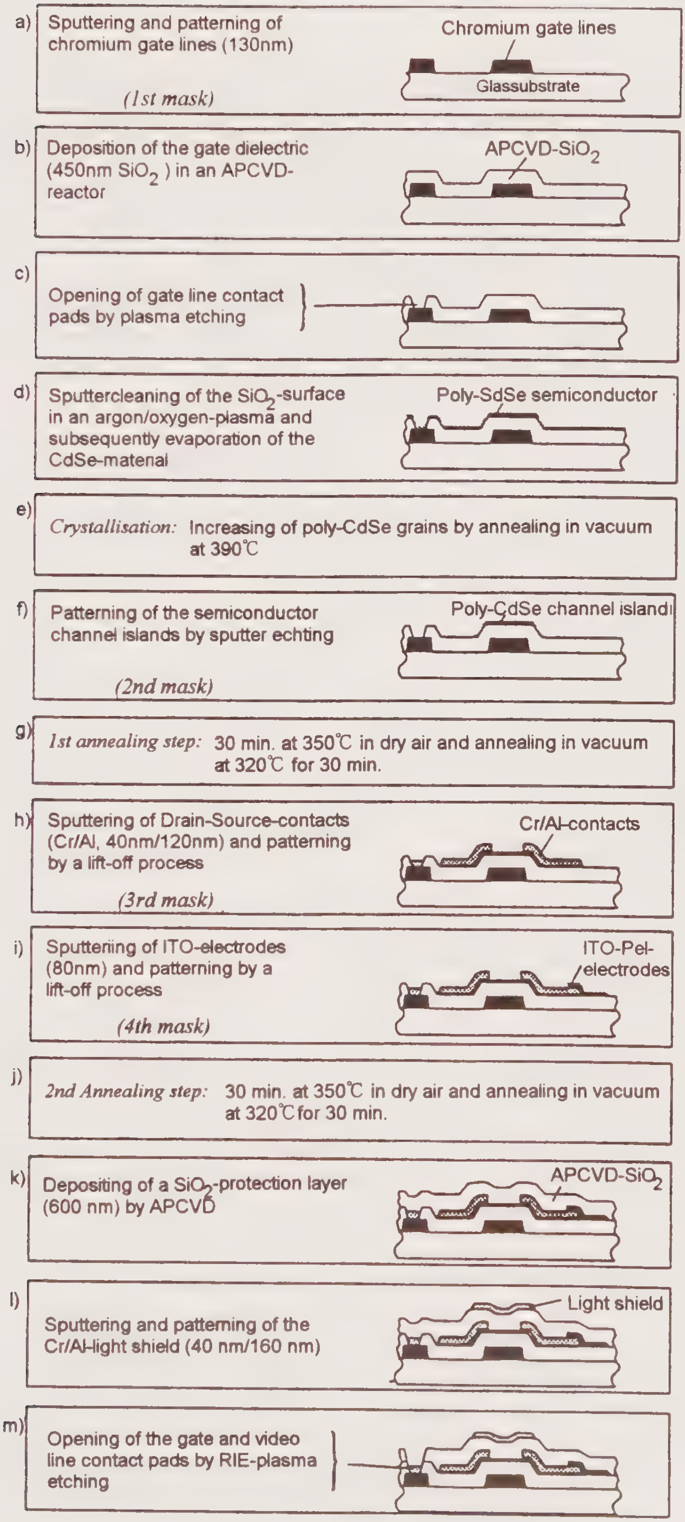

a) Sputtering and patterning of chromium gate lines (130nm)
(1st mask)
Chromium gate lines
Glassubstrate
b) Deposition of the gate dielectric (450nm SiO2) in an APCVD-reactor
APCVD-SiO2
c) Opening of gate line contact pads by plasma etching
d) Sputtercleaning of the SiO2-surface in an argon/oxygen-plasma and subsequently evaporation of the CdSe-material
Poly-SdSe semiconductor
e) Crystallisation: Increasing of poly-CdSe grains by annealing in vacuum at 390℃
f) Patterning of the semiconductor channel islands by sputter echting
(2nd mask)
Poly-CdSe channel island
g) 1st annealing step: 30 min. at 350℃ in dry air and annealing in vacuum at 320℃ for 30 min.
h) Sputtering of Drain-Source-contacts (Cr/Al, 40nm/120nm) and patterning by a lift-off process
(3rd mask)
Cr/Al-contacts
i) Sputteriing of ITO-electrodes (80nm) and patterning by a lift-off process
(4th mask)
ITO-Pel-electrodes
j) 2nd Annealing step: 30 min. at 350℃ in dry air and annealing in vacuum at 320℃ for 30 min.
k) Depositing of a SiO2-protection layer (600 nm) by APCVD
APCVD-SiO2
l) Sputtering and patterning of the Cr/Al-light shield (40 nm/160 nm)
Light shield
m) Opening of the gate and video line contact pads by RIE-plasma etching

FIG. 10.23 CdSe TFT fabrication process.[72]

ration. Following an annealing step at 390°C for 1 hour in vacuum, the CdSe is patterned into channel islands by sputter etch in Ar. There is another annealing step consisting of a 350°C bake in dry air for 30 minutes and a 320°C bake in vacuum. The source and drain contacts are 40-nm Cr and 120-nm Al delineated by a lift-off process. The pixel electrodes are reactively sputtered 80-nm-thick ITO patterned by a lift-off process. After a final annealing step at 350° for 30 minutes in air and at 320°C in vacuum, a 600-nm-thick SiO_2 layer is deposited for passivation. A 40-nm/160-nm Cr/Al light shield is sputter-deposited and etched. Finally, the gate and video line contact pads are exposed by RIE etch. TFTs fabricated with this process yielded mobilities up to 300 cm^2/V-s.

10.6 TFT Array Inspection Testing and Repair

In a TFT array manufacturing line, inspection, testing, and repair have significant importance. In a manufacturing process, inspection and testing can detect defects early in the process. If the defects can not be repaired, scrapping the defective parts early in the process can save costs. Statistics and analyses of defects can provide direction for process developments and yield improvements. Repairing defective arrays can increase yield. There are many types of inspection and testing equipment, including optical inspection,[71,72] voltage imaging,[73] charge sensing,[74] admittance measurement,[75] and e-beam testing.[76]

The optical inspection systems[71,72] use laser scanning, comparing an array image with either a standard image or the image of an adjacent array to detect defects. However, the system can only detect visual anomalies, which may or may not correlate with electrical defects. Optical inspection can be performed at various stages of the fabrication process. One example of an optical inspection system is from KLA, which takes about 1 minute to test one 10.4-inch-diagonal VGA display. Electrical continuity testing uses probes for testing line opens and shorts. One such tester made by Tokyo Cathode has a throughput of about one display per minute. Voltage imaging[73] utilizes a two-dimensional electro-optic modulator to allow direct and noncontact measurement of surface potential and capacitance of a TFT array. By using shorting bars at the edges of the TFT array, electrical signals are fed to the TFT array simulating the display operation. By stepping an electo-optic modulator about 10 μm above the TFT array, circularly polarized light is reflected from the surface of the TFT array. Using the Pockels effects, the surface potential of the TFT array is measured with a resolution of approximately 20 μm and a voltage resolution less than 100 mV to detect line opens, short, and pixel defects. Throughput for a VGA panel for the system is about 3 minutes.

The charge-sensing array tester writes a charge to each pixel and after a holding time transfers the stored charge from the pixel capacitor to an A/D converter.[74] These data are analyzed to detect line and pixel defects. Waveforms used to write and read the pixel charge can be made the same as those used to drive the display. All gate and data lines are contacted separately, therefore a multicontact probe frame is required. Throughput for the tester can exceed one display per minute.

Another type of tester measures the admittance of the TFT array.[75] This tester also needs contacts on each gate and data line. The e-beam testing system[76] scans the TFT array in a vacuum, charging individual pixel electrodes for a short duration and then sensing the secondary electron signals. Analyzing the secondary electron signals can detect various defects in a TFT array. The instrument is even capable of testing the threshold voltage and ON resistance of each TFT. Throughput for this tester is about one display per minute.

Figure 10.24 shows the typical defects in a TFT array. There are opens in data or gate lines and opens in source or drain electrodes. Shorts can be caused by bridging of metal between neighboring components such as ITO pixel electrodes to gate or data lines, or source to drain. Electrical shorts through the insulator can occur in TFTs, in crossovers between gate and data lines, or in storage capacitors. Not all defects can be repaired.

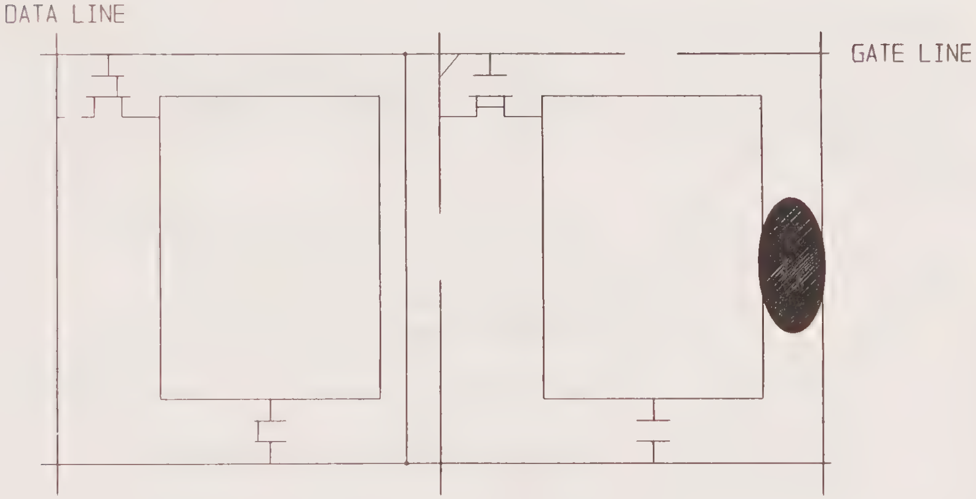


FIG. 10.24 Typical defects in a TFT array.

Opens in bus lines are generally caused by particles embedded in the metal or in the photoresist, causing a hole in the photoresist. Opens in bus lines can be repaired by a laser-assisted deposition process. An example is a Pd deposition process,[76] where a palladium-based metallo-organic film deposited on the TFT substrate by a dispensing system is written by an argon-ion laser depositing a Pd film 8 μm wide and 114 nm thick. The argon-ion laser writes at a 300 μm/s scan rate and produces Pd metal lines with 20-μohm-cm resistivity. Metal bridgings can be removed by cutting and ablating using a pulsed XeCl excimer laser.[77]

10.7 COLOR FILTER FABRICATION PROCESSES

Many types of color filters have been reported in the literature for the color LC display applications. These include dyed gelatin,[78] dyed polyimide,[79] electrodeposition from a dispersion of polymer, dispersing agent, and pigments,[80,81] dispersive pigment color resist,[82] and inorganic dichroic films. A typical fabrication process for dyed gelatin is shown in Fig. 10.25.[78] The gelatin layer has a thickness between 1and 2 μm and is dyed separately three times to obtain the three colors. To enhance the contrast of the cell, a black matrix between the dyed pixels can be included. This black matrix is obtained either by using black-dyed gelatin[83] or pigment-dispersed black organic,[84] or by fabricating the color filter on a substrate with a matrix of black chrome film.[85] Recently, the dispersive-pigment color-resist material has gained wide acceptance because of its better performance in terms of heat and light stability. An example of the fabrication process for the dispersive-pigment color-resist material is shown in Fig. 10.26.[86] The color resist is exposed to UV through a mask and developed like a negative resist. After baking at about 230°C, the resist is cross-linked and hardened. This process is repeated three times for the three colors. The color-resist material has a resolution similar to the negative resist. A geometry down to 5 μm can be obtained.

A pigment-type color filter can also be fabricated by printing, using epoxy resin as the binder.[83] Printing has the advantages of low cost and high throughput. However, the resolution, edge sharpness, and uniformity of the pattern cross section for the printing method are inferior to the photolithographic method. Typical transmissive spectra of a dispersive pigment color filter[86] are shown in

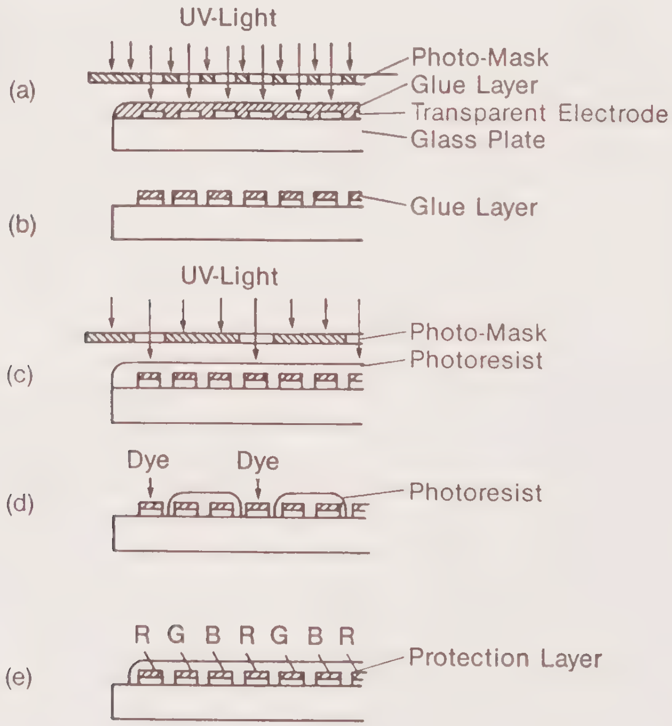


FIG. 10.25 A gelatin color filter fabrication process.[80]

Fig. 10.27. There are four common color pixel arrangements in a color filter as shown in Fig. 10.28.[87] Arrangement 3 is the most popular one for TV applications, while arrangement 4 is considered to be suitable for graphic and alphanumeric applications. Another variation of the arrangement-4 color filter is replacing one of the green pixels with a white pixel to increase the light transmission of the ON state.

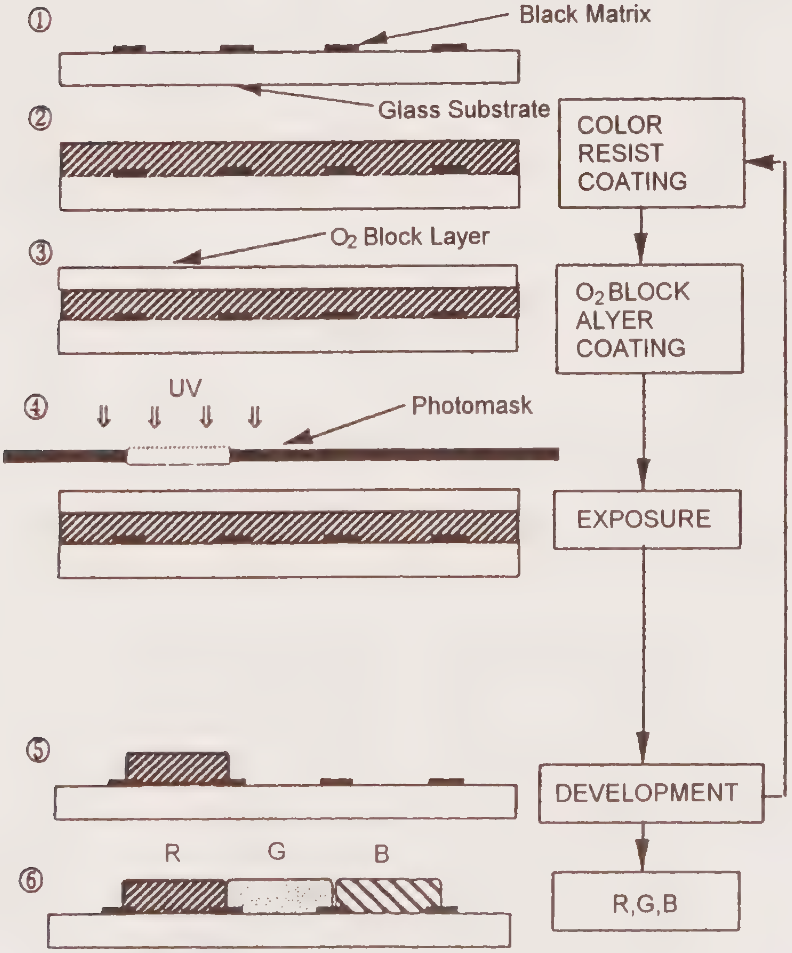

①
Black Matrix
Glass Substrate
②
COLOR RESIST COATING
③
O_2 Block Layer
O_2 BLOCK ALYER COATING
④
UV
Photomask
EXPOSURE
⑤
DEVELOPMENT
⑥
R
G
B
R,G,B

FIG. 10.26 A dispersive pigment color filter fabrication process.[88]

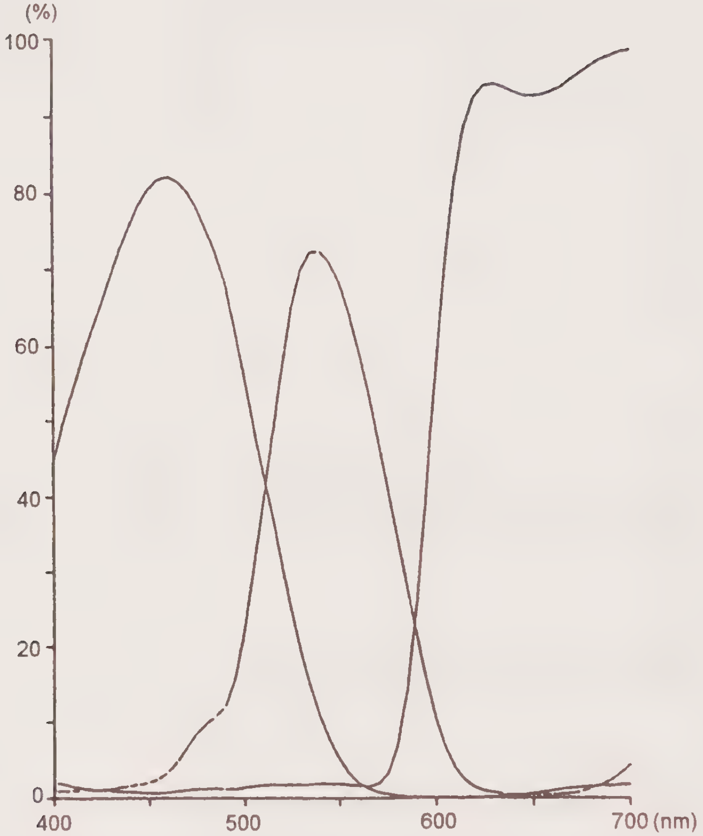
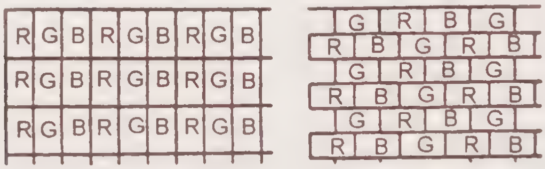


FIG. 10.27 Transmissive spectra of a dispersive pigment color filter.[88]

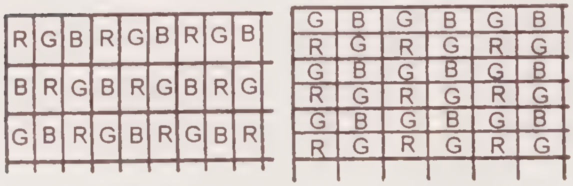


R G B R G B R G B
R G B R G B R G B
R G B R G B R G B

G R B G
R B G R B
G R B G
R B G R B
G R B G
R B G R B

(1) Arrangement-1
(vertical stripe)

(3) Arrangement-3
(triangular)



R G B R G B R G B
B R G B R G B R G
G B R G B R G B R

G B G B G B
R G R G R G
G B G B G B
R G R G R G
G B G B G B
R G R G R G

(2) Arrangement-2

(4) Arrangement-4

FIG. 10.28 Color pixel arrangements in a color filter.

10.8 LC FABRICATION

A cross-sectional view of a TFT-LC cell was shown in Fig. 10.2. The cell consists primarily of two pieces of glass, i.e., one TFT array and one color filter with an LC sandwiched in between. In a typical fabrication process flow, both pieces of glass are coated with a thin layer of polyimide with a thickness ranging from a few tens to 100 nm. Then both plates are rubbed in prearranged directions by a velvet rubbing roller to obtain LC molecule alignments. Next, a narrow strip of epoxy is either dispensed or printed on the borders of the color filter glass to form a ring with a narrow opening. To obtain the proper spacing between the two glass plates, which ranges between 4 and 10 μm, fiberglass spacer rods or spacer beads are dispensed randomly and uniformly on the TFT substrate. Another technique is to delineate a dyed polyimide post on each TFT, which not only maintains the proper spacing but also serves a light-blocking function. The color filter and the TFT array are assembled together with the pixels in the TFT array properly aligned with the pixels in the color filter. The two plates are also arranged to have the surface rubbing directions to be 90° relative to each other so the LC molecules will make a 90° twist going from one plate to the other. After pressing and curing, the assembly is vacuum filled with LC and plugged by quick-setting epoxy.

Despite early attempts to utilize guest-host LC materials[85] to achieve a wider viewing angle, almost all TFT-LC displays now use the TN effect, primarily because of bulk resistivity considerations. In order to achieve the charge storage effect in each pixel in a refresh time frame, it is essential to have, in addition to low TFT OFF current, high resistance in each LC pixel.[30,88] The resistivity of the LC material needs to be greater than 10^{11} ohm-cm at the operating temperature to achieve satisfactory operation. For a cell meeting military specifications, the operating temperature is required to be greater than 85°C. The resistivity of the LC material needs to be greater than 10^{12} ohm-cm at room temperature. The resistivity of the LC material in a cell is also affected by the alignment polyimide material, the buffing material, the cleaning procedure after buffing, and the sealing epoxy. Great care in handling all these steps needs to be exercised to obtain the optimum cell performance.

10.9 OPTICAL PERFORMANCE

At present all TFT-LC devices use the TN effect primarily because of the high bulk resistivity requirement of the TFT-LC circuit. The front and back polarizers of the TN cell can be arranged to be either parallel or perpendicular to each other to obtain either the normally black or normally white effect. The electro-optical measurements of the LC cells with the two arrangements are shown in Fig. 10.29. Usually, cross-polarized cells require higher data voltages than parallel-polarized cells to obtain a saturated dark state. Color TFT-LC devices have been used exclusively in the transmissive mode. To achieve the desired efficiency and brightness in a color LC cell, red, green, and blue fluorescent lamps are used as the backlighting. The spectral distribution[89] of such a lamp is shown in Fig. 10.30.

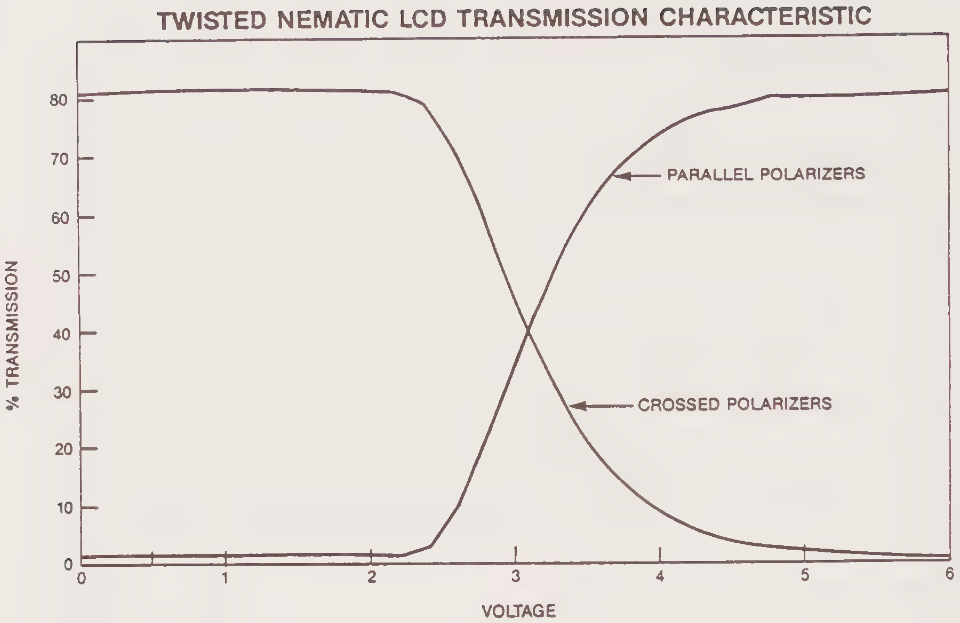


FIG. 10.29 Transmittance vs. driving voltage of TN LC cells with parallel and crossed polarizers.

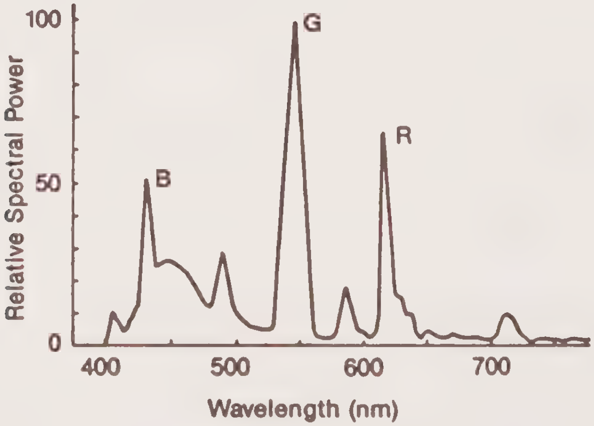


FIG. 10.30 Spectral distribution of an RGB lamp.[92]

In this lamp, specially mixed phosphors are used to generate light emission with spectra matched with the peak transmissions of the color filters to maximize the transmission efficiency.[89–91] The train of light transmission in a typical display is shown in Fig. 10.31. The light emitted from fluorescent lamps passes through a diffuser, polarizer, TFT array, color filter, analyzer, and antireflection layer to reach the viewer. To maximize the transmission efficiency, each layer in the light path needs to be optimized. The key factors are the lamp design, the color filter transmission, and the aperture ratio in the cell design. A typical transmission efficiency of 3–5% after the diffuser is reported.

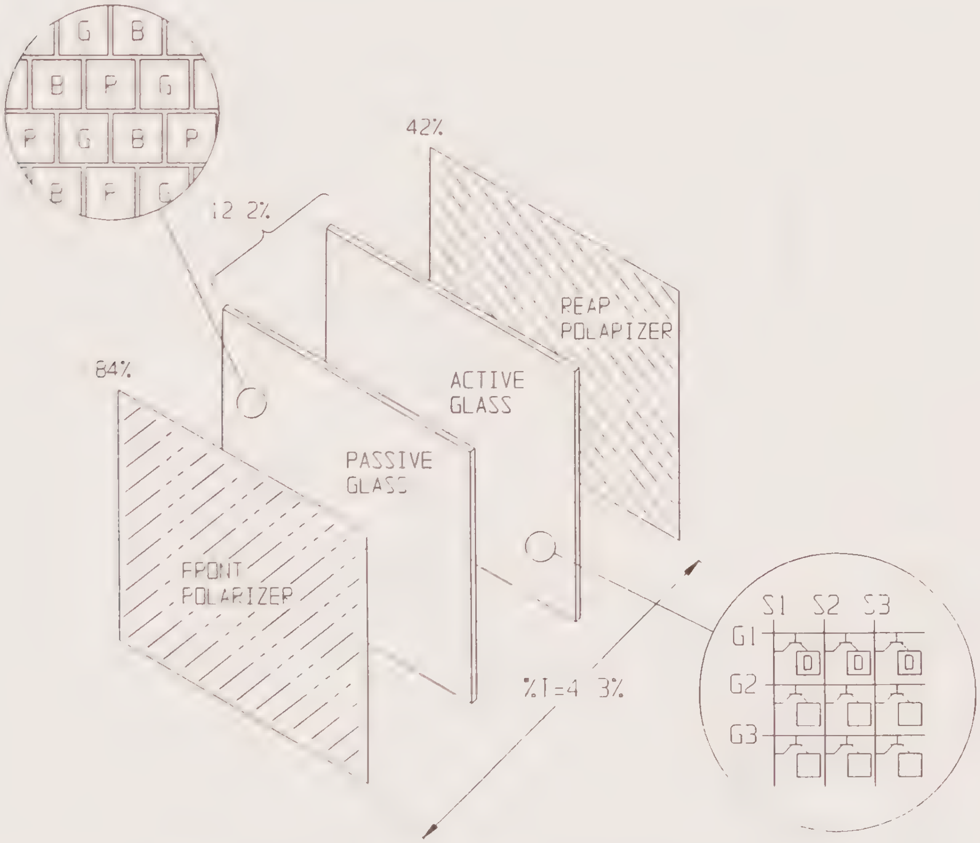


FIG. 10.31 Light transmission in a TFT-LCD.

There are both advantages and disadvantages to parallel and crossed polarizers. Figs. 10.32(a) and (b)[92] show the contrast ratios vs. Δnd for cells with parallel and crossed polarizers, respectively, where Δn is the birefringence of the LC material and d is the cell thickness in micrometers. As shown in the figures, the contrast of parallel-polarized cells has a stronger dependence on Δnd than cross-polarized cells. Parallel-polarized cells require critical cell thickness control to obtain a very dark state and an acceptable dark state uniformity. Parallel-polarized cells also have more severe degradation of the dark state at elevated temperatures when Δn is reduced.

Because of the anisotropy of the polarizers and the LC material, the appearance of LC cells is angle dependent. Figures 10.33 and 10.34 show the transmittance vs. driving voltage curves for different viewing angles for cells with parallel and crossed polarizers, respectively.[93] There are significant top-to-bottom brightness variations for both parallel- and cross-polarizer modes. The left-to-right brightness variation is less for the cross-polarizer mode than for the parallel-polarizer mode.

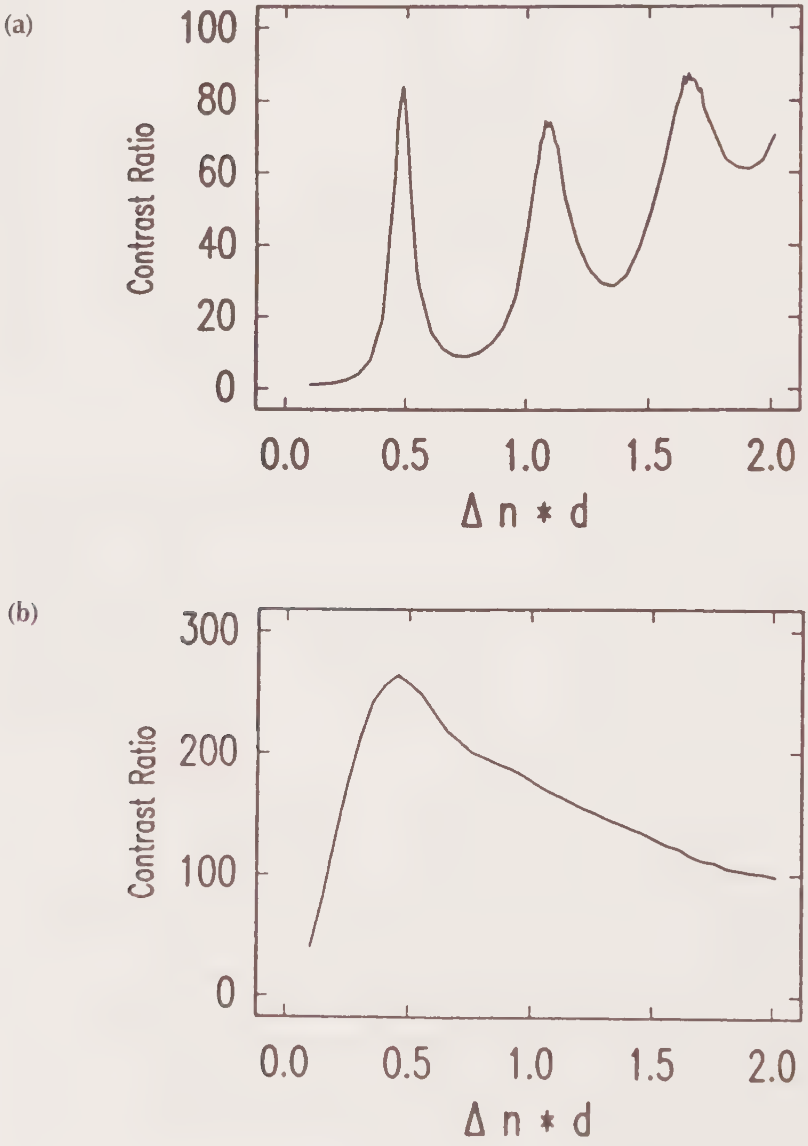


FIG. 10.32 (a) Contrast ratios vs. nd for cells with parallel polarizers[95] (b) Contrast ratios vs. nd for cells with crossed polarizers[95]

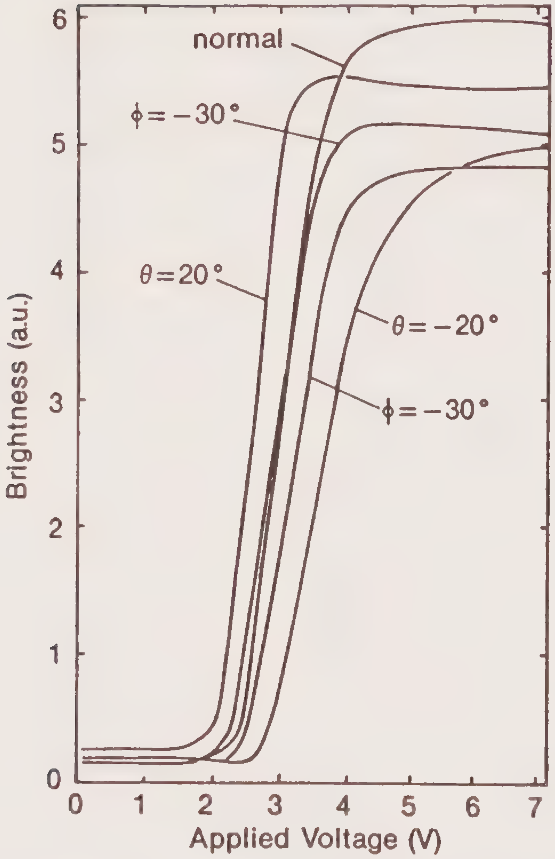


FIG. 10.33 Cell brightness vs. driving voltage for different viewing angles for a cell with parallel polarizers. θ is for vertical and ϕ for horizontal viewing angles.[96]

There are some color shifts of the display vs. viewing angles also. The comparison of the black colors of the dark states of the parallel- and crossed-polarizer cells is shown in Fig. 10.35.[94] As indicated in this figure, there are more color shifts of the dark state in the parallel-polarizer mode than in the crossed-polarizer mode.

Two reported display panels are used here to exemplify the TFT-LC cell performance. One is a crossed-polarizer panel,[95] the other a parallel-polarizer panel.[96] The contrast ratio dependence of viewing angles of the crossed-polarizer panel is shown in Fig. 10.36. A contrast ratio of over 200:1 was obtained. The color coordinates of the cell back-illuminated by a fluorescent lamp are shown in Fig. 10.37. The contrast ratio dependence of viewing angle and the color coordinates of the parallel-polarizer cell are shown in Figs. 10.38 and 10.39,[83] respectively. A contrast ratio of over 50:1 was reported.

Significant progress has been made in improving the viewing-angle performance of TFT-LCDs. One approach uses compensators. By inserting compensators

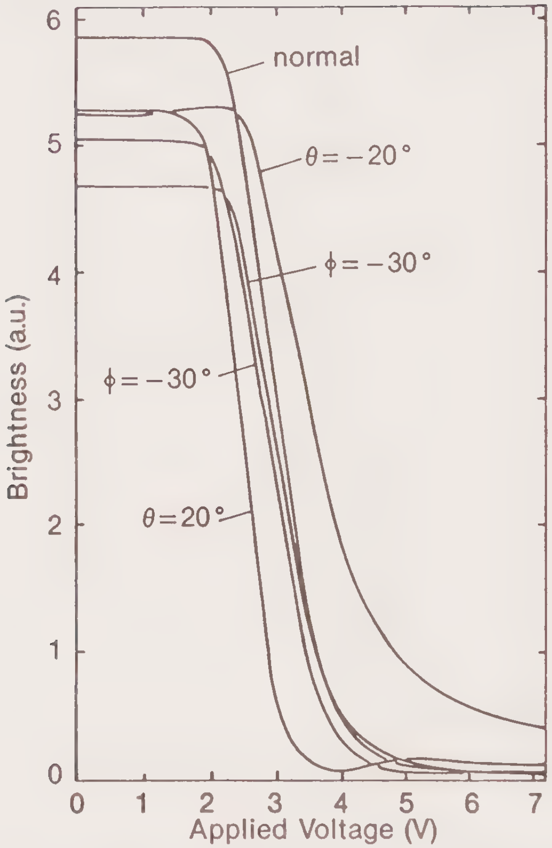


FIG. 10.34 Cell brightness vs. driving voltages for different viewing angels for a cell with crossed polarizers. θ is for verticle and ϕ for horizontal viewing angles.[96]

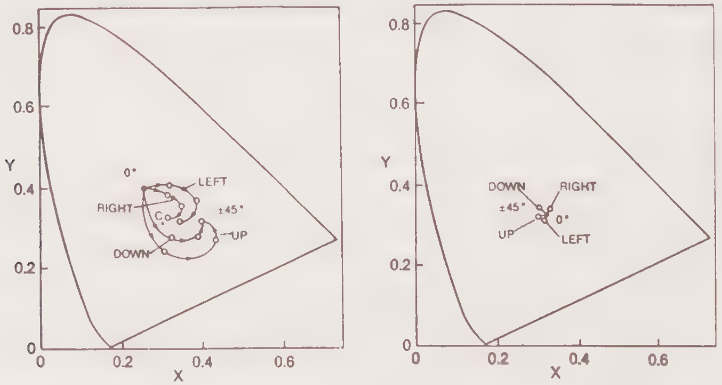


FIG. 10.35 The dark state color variations vs. viewing angles of TFT-LC cells on the CIE chromaticity diagrams with (a) parallel and (b) crossed polarizers.[97]

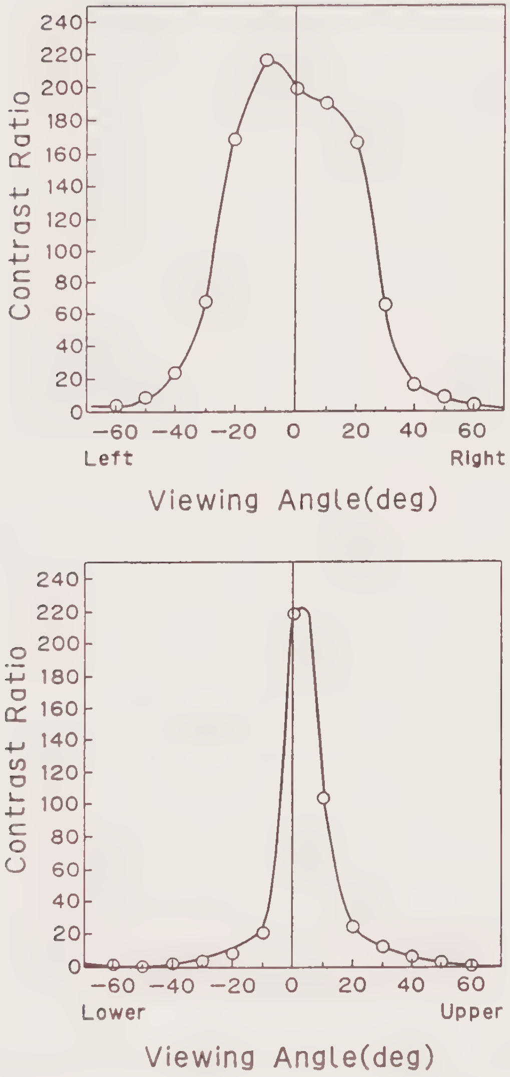


FIG. 10.36 Viewing angle dependence of contrast ratio for a crossed-polarizer TFT-LC cell. Top: horizontal direction. Bottom: vertical direction.[98]

between the polarizer and analyzer adjacent to the LC cell, the contrast ratio variation vs. viewing angle can be reduced.[96]

For normally white (NW) displays, negative-Δn compensators[97] can increase the horizontal and vertical viewing angles by more than 10°. For normally black (NB) displays, compensators with a positive Δn are used.[98] The improved performance is shown in Fig. 10.40.[99] The regions with a contrast ratio greater than 30:1 are extended to greater than 60° horizontal.

One of the major shortcomings of the NW display is its asymmetric and limited viewing capability and gray-scale inversion in the vertical direction. One method

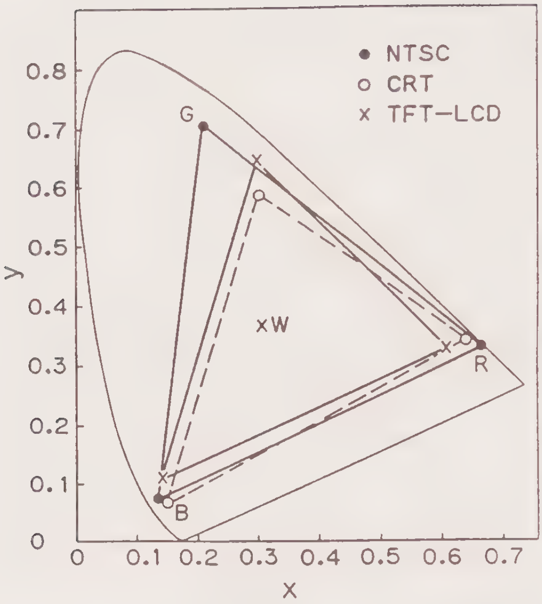


FIG. 10.37 CIE color coordinates of a crossed-polarizer TFT-LC cell back-illuminated by fluorescent lamps.[98]

to improve the performance is to divide each pixel into two or even multidomains of alignments.[100–102] The opposite LC orientations in each LC pixel compensate each other and achieve symmetry in the vertical direction. Figs. 10.41(a) and (b) compare gray-level performance of a regular TN panel with a domain-divided LC panel in the vertical direction.[101] Symmetrical contrast and noninversion in gray scale have been achieved in the domain-divided cell.

Another recent major improvement in viewing-angle performance is the in-plane-switching (IPS) mode,[103,104] where homogeneously aligned LC molecules are rotated by electric fileds parallel to the substrate surface. With a cross-polarizer arrangement, the display appears dark without electric field. Using an LC material with a negative dielectric constant, interdigital electrodes on the TFT array substrate provide the electric field to rotate the LC molecules and turn the display into transmissive. Wide viewing angles greater than 70° in both up-down and left-right directions with little gray-level inversion have been obtained.

TFT-LCDs have very good viewability under high-brightness ambient conditions. Low reflectivity of TFT-LCD for automotive applications was reported with a reflectance of 1.1%.[105] As shown in Fig. 10.42, the total low reflectance was obtained primarily by optimizing the front antireflection coating and the low-reflection black matrix. For military avionic applications, a contrast ratio greater than 10:1 in a wide viewing envelope under 10,000 foot-candle illumination was reported.[90]

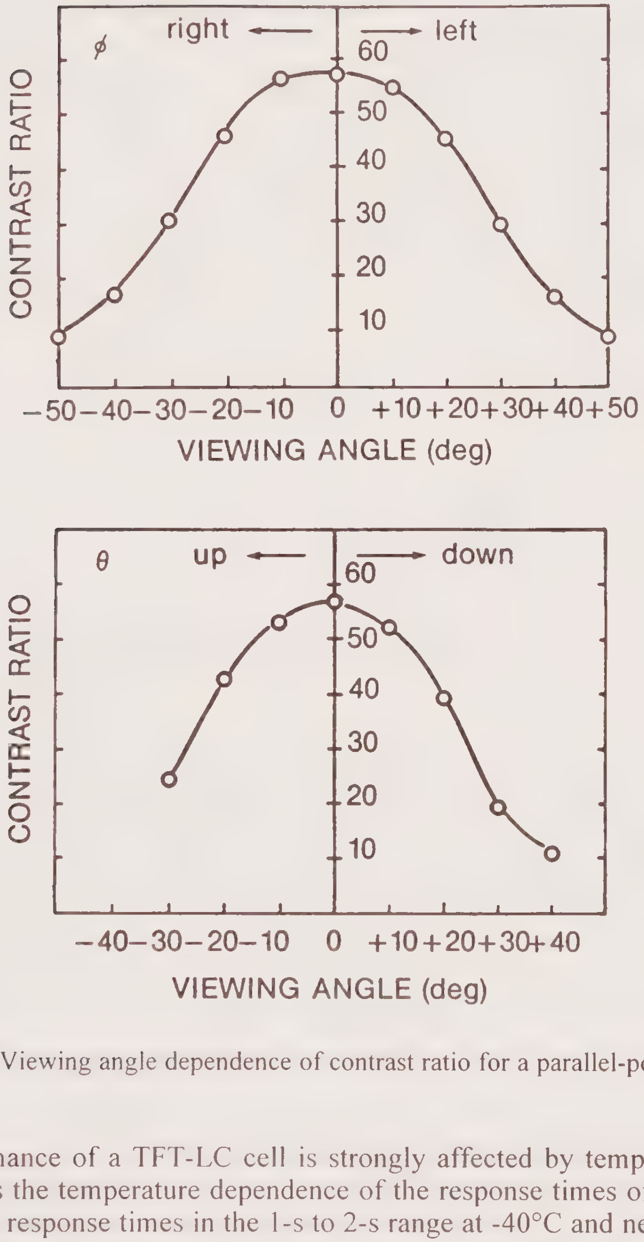


FIG. 10.38 Viewing angle dependence of contrast ratio for a parallel-polarizer TFT-LC cell.[85]

The performance of a TFT-LC cell is strongly affected by temperature. Figure 10.43 shows the temperature dependence of the response times of an LC cell.[106] The cell has response times in the 1-s to 2-s range at -40°C and needs a heater to obtain satisfactory operation at cold temperatures. The ON and OFF currents of the TFT,[107] and the resistivity, threshold voltage (decreasing by about 0.1 V per 10°C increase), and birefringence (decreasing by about 0.003 per 10°C increase) of the LC material, all change significantly at elevated temperatures, which cause a strong temperature dependence of important cell characteristics such as contrast ratio, viewing angle, and color gamut. Nevertheless, operation of TFT-LC cells between –40°C and –85°C has been demonstrated.

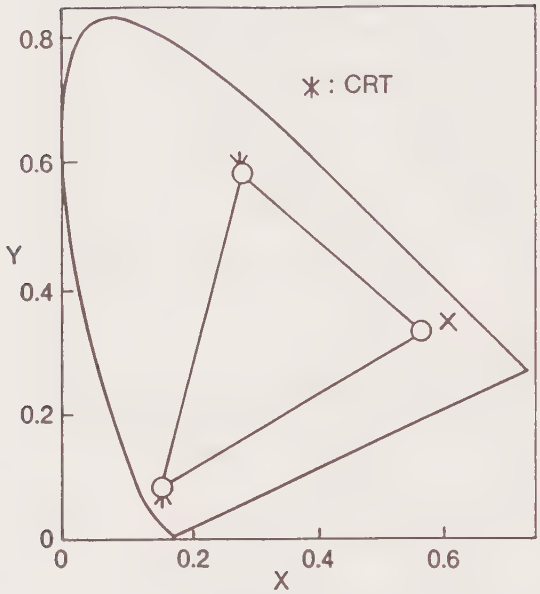


FIG. 10.39 CIE color coordinates of a parallel-polarizer TFT-LC cell.[85]

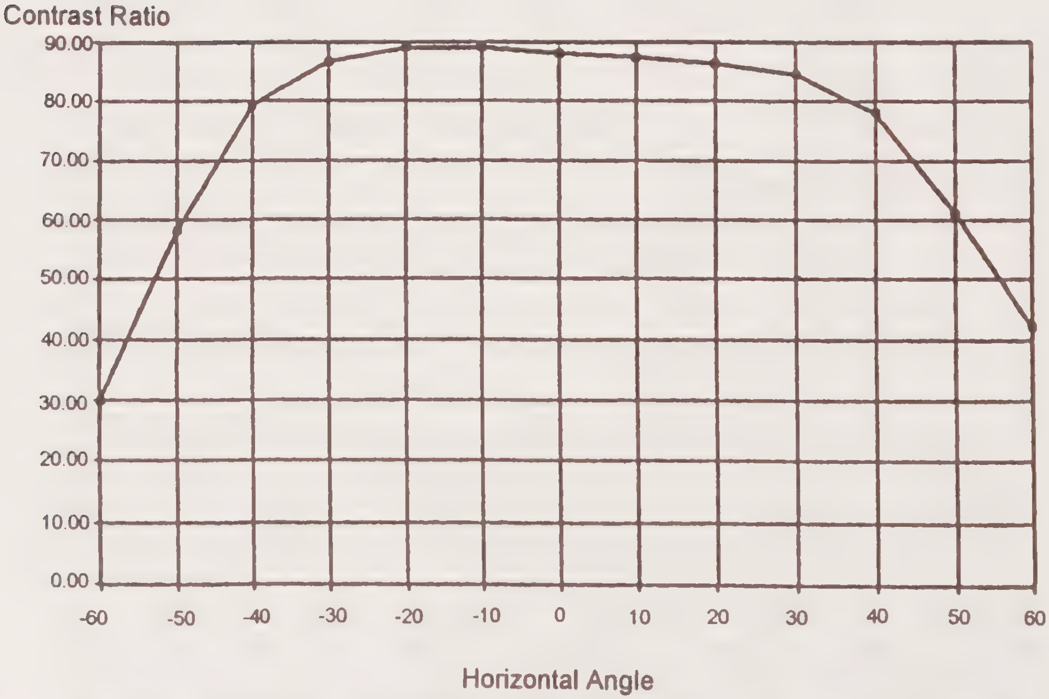
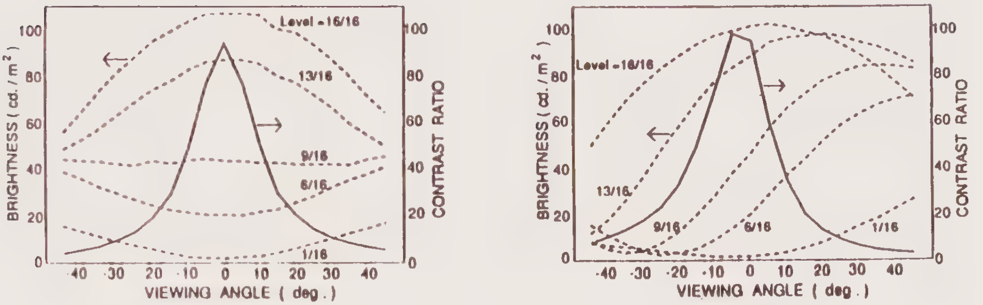


FIG. 10.40 Contrast ratio vs. horizontal viewing angles for a compensated NB display.[36]



(a) DDTN PANEL (b) TN PANEL

FIG. 10.41 Viewing angle performance of (a) a domain-divided TN cell and (b) regular TN cell.[101]

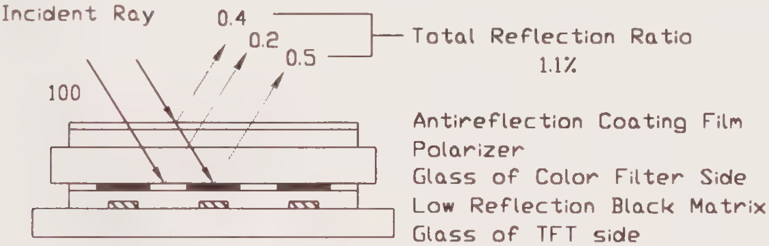

Incident Ray
100
0.4
0.2
0.5
Total Reflection Ratio
1.1%
Antireflection Coating Film
Polarizer
Glass of Color Filter Side
Low Reflection Black Matrix
Glass of TFT side

FIG. 10.42 Reflections in a TFT-LCD.[105]

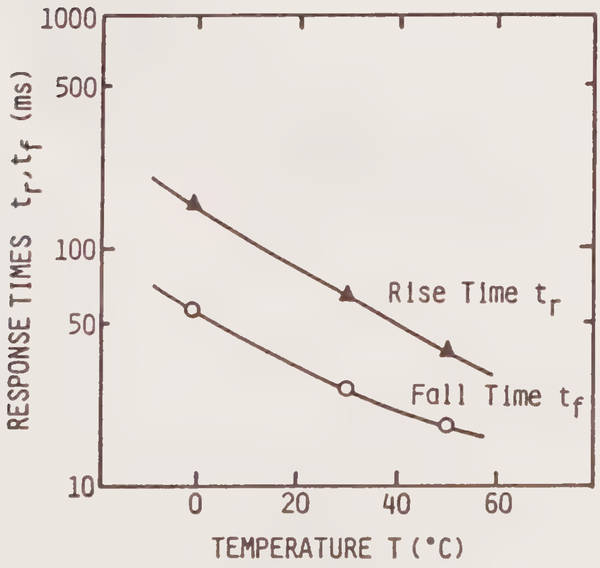


FIG. 10.43 Temperature dependence of the response time of an LC cell.[106]

10.10 DISPLAY APPLICATIONS AND STATUS

In the past few years, the development of TFT-LC devices has been progressing at an explosive pace. The sales of TFT-LC devices, as shown in Fig. 10.44, are predicted to be over \$10 billion by the year 2000. Figures 10.45 and 10.46[108] summarize recent prototype development trends in display diagonal sizes and resolution, respectively. The displays are getting larger and denser every year. The largest display size has been expanded to a 21-inch-diagonal,[1] the highest resolution to 1000 lines per square inch, and the most display pixels to 6 million.[42]

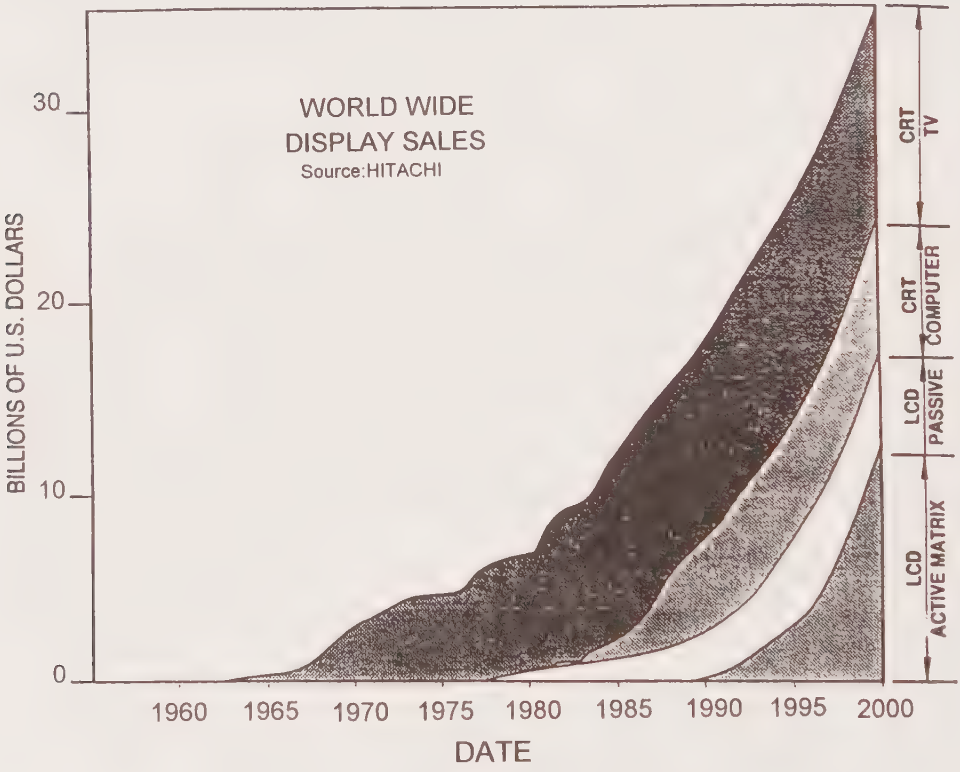


FIG. 10.44 World-wide display sales forecast.

Currently, TFT-LC devices have been utilized in applications such as portable TVs, games, viewfinders, camcorders, virtual reality displays, automotive global positioning systems, projection TVs, computer displays, instrumentation displays, and military avionic displays. For the projection-TV application, high-definition TVs with up to 1840×1035 pixels resolution[48] have been demonstrated. For computer displays, 12.1-inch notebook PCs and large-area high-information-content displays up to 1280×3×1024 pixels for engineering work station (EWS) applications[109] are being developed rapidly. In almost all of these applications, the color TFT-LC devices have adequate color gamut, gray-scale capability, re-

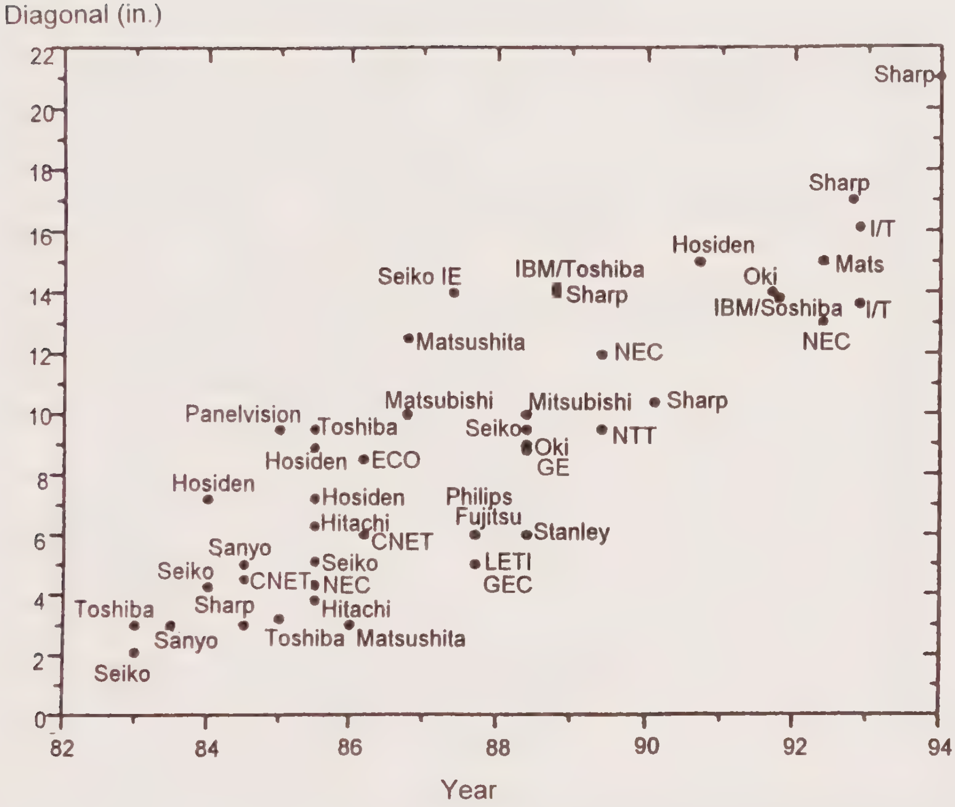


FIG. 10.45 TFT-LCDs trend of increasing size with time.[108]

sponse time, resolution, power consumption, and high ambient viewability to compete with CRTs. The areas where improvements are required are manufacturing cost, large-size availability, and viewing angle.

At this moment, the price of TFT-LCDs is quickly approaching the goal of $10 per square inch. The cost can be further lowered through throughput improvements of fabrication equipment and reductions in component costs. The developments of integrating driver circuits on the same substrate as the TFT array by poly-Si devices or using the chip-on-glass technology can reduce the cost of driver electronics.

The fabrication of large-area devices up to 40-inch-diagonal is primarily limited by the availability of processing equipment. Besides the requirement for development of some high-conductivity metal layers for the row and column lines, there should be no fundamental limitation for expansion into a large area.

Viewing-angle dependence is an inherent quality of TN LC displays. The viewing-angle dependence of TFT-LC devices can be further improved through compensators and multidomain process development. Also, some other LC effects

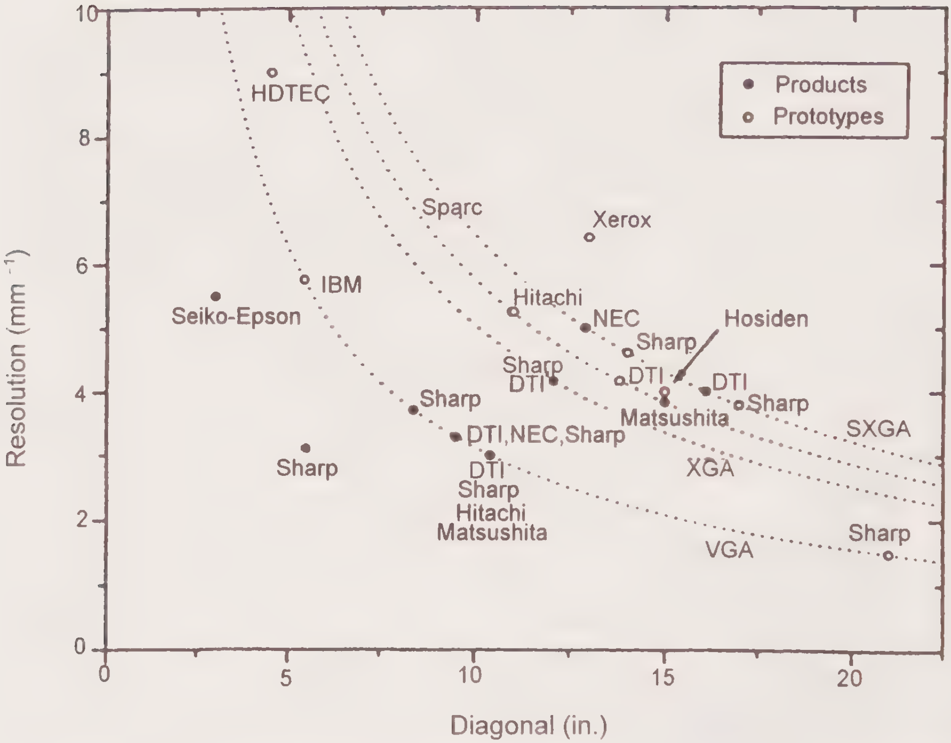


FIG. 10.46 Recent TFT-LCD prototypes plotted by resolution and diagonal size.[108] The dotted lines denote fixed content, e.g. VGA.

incorporated with TFT addressing will be developed in the future that can drastically improve the viewing angle performance.

10.11 CONCLUSION

In the past few years, the TFT-LC device has made tremendous progress in manufacturing technology and optical performance, and has developed as a strong contender to challenge the CRT. In many areas, such as weight, power consumption, viewability under strong ambient light, response time, color gamut, and cell brightness, TFT-LC devices have performance equivalent or superior to CRTs. TFT-LC devices still need to improve on their manufacturing cost, availability in a large size, and viewing angle to be widely accepted by users.

REFERENCES

1. M. Hijikigawa and H. Take, *SID '95 Digest*, p. 147 (1995).
2. F. Cuomo et al., *SID '95 Digest*, p. 77 (1995).
3. D. R. Baraff et al., *1980 Biennial Display Res. Conf.*, p. 109 (1980).
4. S. Togashi et al., *SID '84 Digest*, p. 324 (1984).
5. N. Szydia et al., *Japan Display '83 Proc.*, p. 416 (1983).

6. Z. Yaniv et al., *SID '88 Digest*, p. 420 (1988).
7. M. Toyama, *SID '87 Digest*, p. 155 (1987).
8. J. E. Lilienfeld, U.S. Patent 1900 018, Mar. 1933.
9. O. Heil, British Patent 439 457, Dec. 1935 (German patent date 1934).
10. P. K.Weimer, *Proc. IRE*, **50**, p. 1462 (1962).
11. R. Zuleeg, *Solid-State Electron.*, **6**, p. 645 (1963).
12. H. Koelmans and H.C. de Graaff, *Solid-State Electron.*, **10**, p. 997 (1967).
13. C. A. Neugebauer et al., *Thin Solid Films*, **2**, p. 57 (1968).
14. B. J. Lechner, *Proc. IEEE*, **59**, p. 1566 (1971).
15. A. G. Fisher et al., *Conf. Rec. IEEE Conf. on Display Devices*, p. 64 (1972).
16. A. G. Fisher, U.S. Patent 3,840,695, Oct. 1974.
17. T. P. Brody et al., *IEEE Trans. Electron. Dev.*, **ED-20**, p. 995 (1973).
18. T. P. Brody et al., IEDM, Washington D.C., (1972).
19, T. P. Brody et al., IEDM, Washington D.C., p. 200 (1975).
20. T. P. Brody el al., *SID '74 Digest*, p. 166 (1974).
21. F. C. Luo et al., *SID '78 Digest*, p. 94 (1978).
22. F. C. Luo and W. Hester, *IEEE Trans. Electron. Dev.*, 27, p. 223 (1980).
23. F. C. Luo and D. Hoesly, *SID '82 Digest*, p. 46 (1982).
24. F. C. Luo and K. Richardson, *SID '83 Digest*, p. 184 (1983).
25. F. C. Luo et al., *SID '85 Digest*, p. 286 (1985).
26. S. Weber, *Electronics*, May, p. 97 (1989).
27. F. C. Luo, P. Pultorak, and E. Freeman, *IEEE Trans. Electron. Dev.*, **30**, p. 202 (1983).
28. Y. Baron et al., *Proc. Japan Display '86*, p. 68 (1986).
29. T. Shimasaki et al., *Proc. Japan Display '88*, p. 78 (1988).
30. T. L. Credelle, *IDRC '88 Conf. Record*, p. 208 (1988).
31. S. Hott et al., *SID '86 Digest*, p. 296 (1986).
32. T. Yanagisawa, K. Kasahara, and M. Kajimura, *Proc. Japan Display '86*, p. 192 (1986).
33. M. J. Powell, *Proc. Eurodisplay '87*, p. 63 (1987).
34. W. E. Howard, P.M. Alt, and R.L. Wisnieff, *IDRC '88 Conf. Record*, p. 230 (1988).
35. D. Castleberry and G. Possin, *SID '88 Digest*, p. 232 (1988).
36. T. Katagishi et al., *SID '86 Digest*, p. 285 (1986).
37. L. T. Lipton and N.J. Koda, *SID '73 Digest*, p. 46 (1973).
38. M. Dobler et al., “Fabrication of CdSe TFTs with high mobility of 300 cm^2/V-s for a 140,000 pel AMLCD,” 3rd Int. CdSe Workshop.
39. J. F. Farrell et al., *Japan Display '92*, p. 885 (1992).
40. K. Schleupen et al., *SID '95 Digest*, p. 85 (1995).
41. P. G. LeComber et al., *Electron. Lett.*, **15**, p. 179 (1979).
42. R. Martin et al., *SID '97 Digest*, p. 7 (1997).
43. Y. Okita et al., *SID '91 Digest*, p. 411 (1991).
44. K. Takeuchi et al., *SID '91 Digest*, p. 415 (1991).

45. R. G. Stewart et al., *SID '95 Digest*, p. 89 (1995).
46. H. Lebrunn et al., *SID '95 Digest*, p. 493 (1995).
47. S. Higashi et al., *SID '95 Digest*, p. 81 (1995).
48. M. Kobayashi et al., *SID '94 Digest*, p. 75 (1994).
49. I-W. Wu et al., *SID '92 Digest*, p. 615 (1992).
50. T. Little et al., *IDRC '91 Conf. Record*, p. 219 (1991).
51. H. Oshima et al., *SID '93 Digest*, p. 387 (1993).
52. J. B. Youn et al., *SID '95 Digest*, p. 321 (1995).
53. K. R. Sarma et al., *SID '94 Digest*, p. 419 (1994).
54. J. P. Salerno et al., *SID '92 Digest*, p. 63 (1992).
55. F. Funada et al., *SID '86 Digest*, p. 293 (1986).
56. T. Nagayasu et al., *IDRC '88 Conf. Record*, p. 56 (1988).
57. T. Tsukada, *Japan Display '92*, p. 325 (1992).
58. M. Katayama et al., *IDRC '91Conf. Record*, p. 243 (1991).
59. S. Morozumi, "Issues in Manufacturing Active Matrix LCDs", SID Seminar lecture notes, p. F3/3, 1992.
60. M. Akiyama et al., *SID '93 Digest*, p. 887 (1993).
61. T. Sakai et al., *Proc. Japan Display '86*, post-deadline 3 (1986).
62. T. Sunata et al., *Proc. SID*, **27**, p. 229 (1986).
63. H. Moriyama et al., *SID '89 Digest*, p. 144 (1989).
64. M. J. Powell., *Appl. Phys. Lett.*, **43**, p. 597 (1983).
65. K. Suzuki et al., *IDRC '85 Conf. Record*, p. 14 (1985).
66. F. R. Libsch and J. Kanicki, *SID '93 Digest*, p. 455 (1993).
67. M. Matsuo et al., *SID '94 Digest*, p. 87 (1994).
68. I. Yudasaka et al., *Proc. Japan Display '92*, p. 451 (1992).
69. K. Nakazawa et al., *SID '90 Digest*, p. 311 (1990).
70. E. Lueder, *IDRC '94 Conf. Record*, p. 30 (1994).
71. D. Hendrick et al., *SID '92 Digest*, p. 46 (1992).
72. J. H. Chen and H. K. Tsai, *SID '94 Digest*, p. 444 (1994).
73. F. Henley and G. Addiego, *SID '91 Digest*, p. 687 (1991).
74. S. Kimura et al., *SID '92 Digest*, p. 628 (1992).
75. H. P. Hall and P.R. Pilotte, *SID '91 Digest*, p. 682 (1991).
76. M. Brunner et al., *Proc. EuroDisplay '93*, p. 387 (1993).
77. K. Shen et al., *SID '92 Digest*, p. 636 (1992).
78. T. Uchida et al., IEEE Trans. Electron. Dev, ED-30, p. 503 (1983).
79. W. I. Latham et al., SID '87 Digest, p. 379 (1987).
80. M. Miyamura "Manufacture of Color Filter", Jap. Pat. Appl. 59-78312.
81. H. Kamamura, SID '84 Digest, p. 215 (1984).
82. T. Sugiura et al., SID '92 Digest, p. 21 (1992).
83. S. Hota et al., SID '86 Digest, p. 296 (1986).
84. C. M. Hosler et al., SID '95 Digest, p. 446 (1995).
85. T. Sunata et al., IDRC '85 Conf. Record, p. 24 (1985).
86. M. Tani and T. Sugira, IDRC '94 Conf. Record, p. 103 (1994).

87. S. Tsuruta et al., IDRC '85 Conf. Record, p. 24 (1985).
88. J. Varney and K.E. Crary, SID '89 Digest, p. 163 (1989).
89. F. Inoue et al., SID '88 Digest, p. 318 (1988).
90. E. Haim et al., AIAA/IEEE, *Proc. 8th Digital Avionics Systems Conf.*, p. 564 (1988).
91. W. W. Piper, *SID '88 Digest*, p. 326 (1988).
92. A. Lien and S. Suzuki, *Proc. Japan Display '89*, p. 286 (1989).
93. Y. Miyata et al., *SID '89 Digest*, p. 155 (1989).
94. M. Katayama et al., *SID '88 Digest*, p. 310 (1988).
95. H. Maeda et al., *SID '92 Digest*, p. 47 (1992).
96. S. Kaneko et al., *SID '93 Digest*, p. 265 (1993).
97. H. Mori et al., *SID '97 Digest*, p. 941 (1997).
98. N. Yamagishi, H. Watanabe, and K. Yokoyama, *Proc. Japan Display '89*, p. 316 (1989)
99. S. Thomsen, G. Fye, and A. Abileah, *Proc. SPIE*, **2219**, p. 130-139 (1994)
100. K. H. Yang, *IDRC '91 Conf. Record*, p. 68 (1991).
101. Y. Koike et al., *SID '92 Digest*, p. 798 (1992).
102. K. Takatori et al., *Japan Display '92 Proc.*, p. 591 (1992).
M. Oh-e et al., *Asia Display '95 Proc.*, p. 577 (1995).
M. Ohta et al., *Asia Display '95 Proc.*, p. 707 (1995).
105. H. Yoshida et al., *SID '94 Digest*, p. 27 (1994).
106. J. Sasuga et al., *Proc. Japan Display '89*, p. 510 (1989).
107. F. C. Luo et al., *SID '88 Digest*, p. 235 (1988).
108. W. E. Howard, *J. Soc. Inf. Display*, **3**, p. 127 (1995).
109. Y. Numano et al., *SID '97 Digest*, p. 11

CHAPTER 11

Microsensors, Actuators, MEMS, and Electronics for Smart Structures

Vijay K. Varadan
Vasundara V. Varadan
The Pennsylvania State University

CONTENTS

11.1 Introduction / 619

11.2 Multifunctional Smart Materials / 623

11.3 Microelectromechanical Systems / 624
- 11.3.1 Micromachining and microfabrication / 625
  - 11.3.1.1 Bulk micromachining / 625
  - 11.3.1.2 Surface micromachining / 626
  - 11.3.1.3 Stereo lithography / 627

11.4 Smart Electronics and Conformal Antennas / 630
- 11.4.1 Application-specific integrated circuits / 633
  - 11.4.1.1 Accuracy of the hybrid technique / 635
- 11.4.2 Wireless telemetry systems / 642

11.5 Microsensors for Smart Structures / 645
- 11.5.1 Interdigital transducer microsensors / 646
- 11.5.2 Microaccelerometer / 651
- 11.5.3 Programmable accelerometer / 652
  - 11.5.3.1 Fabrication method / 659
- 11.5.4 Suspension design / 659
  - 11.5.4.1 Adhesion / 659
  - 11.5.4.2 Adhesion prevention / 660
- 11.5.5 Wafer bonding / 660
- 11.5.6 Other design considerations / 660
- 11.5.7 Acoustic attenuation / 661
- 11.5.8 Advantages of the design presented / 661

11.6 Sensors for Health Monitoring / 662

11.7 Applications to Smart Structures / 664
- 11.7.1 Deflection and strain of flex-type beam structures / 664
  - 11.7.1.1 Experimental setup and results / 666
- 11.7.2 Dynamic strain / 668
- 11.7.3 Local and global sensing and control of vibration and drag in aircraft / 671
- 11.7.4 "Smart Plate" for noise reduction using MEMS technology / 676
- 11.7.5 Health monitoring of aircraft structures / 681

Conclusion / 684

Acknowledgments / 685

References / 685

11.1 INTRODUCTION

In this chapter, the integration of multifunctional smart materials, microelectromechanical systems (MEMS), and smart electronics is presented with examples covering some civil and aerospace structures and devices. Some applications include systems for noise suppression in aircraft cabin using "smart wall paper," drag sensing and reduction in aircraft, sensing and control of ice formation and deicing on aircraft, remote measurement of tip deflection of helicopters using MEMS sensor and electronically tunable antenna, smart skin antennas for communication between sensors and actuators (telemetry devices), and health monitoring of rotorcrafts using MEMS sensors along with fiber optics sensor and remote antenna systems.

The term *multifunctional smart material* refers to a material that senses changes in its environment and makes a useful or optimal response by changing its material properties, geometry, or mechanical or electromagnetic response. Both the sensor and actuator functions with the appropriate feedback must be integrated and these comprise the "brain" of the material. The materials belonging to this category may include a range of artificial materials from optically active or chiral polymers to multifunctional polymers, and ferroelectric or other active ceramics. Advanced smart polymeric composites are a new generation of multifunctional polymers, piezopolymers, and conducting polymers that must be distinguished from piezoelectric polymers such as PVDF that are in current use. Piezopolymers involve the integration of polymers and nanoceramic particles by chemical bonding as side groups on a polymer backbone. The nanoparticles have active surfaces or functional groups that can bond with the polymer chain. The ceramic side groups provide the piezoelectric function in the polymer, while the backbone provides mechanical strength, structural integrity, electrical conductivity, etc.

MEMS refers to microsensors and actuators that sense their environment and can react to changes in that environment with the use of microcircuit control. In addition to including conventional microelectronics packaging, these systems integrate antenna structures for command signals into microelectromechanical structures for desired sensing and actuating functions. The systems may also need a micropower supply, microrelay, and microsignal processing units. Microcomponents make the systems faster, more reliable, cheaper, and capable of incorporating more complex functions.

Micromachining has become the fundamental technology for the fabrication of microelectromechanical devices and, in particular, miniature sensors and actuators. Silicon micromachining is the most mature of the micromachining technologies, and it allows for the fabrication of MEMS that have dimensions in the submillimeter range. It refers to fashioning microscopic mechanical parts out of silicon substrate or on a silicon substrate, making the structures three dimensional and bringing new principles to the designers. Employing materials such as crystalline silicon, polycrystalline silicon, and silicon nitride, a variety of mechanical microstructures including beams, diaphragms, grooves, orifices, springs, gears, suspensions, and other complex mechanical structures has been conceived.[1–5]

Many microdevices can also be fabricated using semiconductor processing technologies or stereo lithography on the polymeric multifunctional sheets mentioned above. With proper doping, a semiconducting polymer structure can be synthesized, using stereo lithography. It is now possible to make three-dimensional microstructures of high aspect ratio using this technique. Ikuta et al.[6] demonstrated that a three-dimensional microstructure of polymers and metal is feasible using an "IH process" (integrated hardened polymer stereo lithography). Using a UV light source, XYZ-stage, shutter, lens, and microcomputer, they have shown that microdevices such as springs, venous valves, and electrostatic microactuators can be fabricated. In case of difficulty with the polymeric materials, some of these devices can be micromachined in silicon and the system architecture obtained by photoforming or hybrid processing.[6–12] The photoforming or photofabrication is an optical method such as stereo lithography, photomask layering process, or IH process that involves solidification of photochemical resin by light exposure. Takagi and Nakajima[12] proposed new concepts of "combined architecture" and "glue mechanism" using the photoforming process to fabricate complicated structures by combining components, each of them made by its best fabrication process. Batch processing of such devices thus seems feasible.

The combined architecture may also result in sheets of smart skins with integrated sensors and actuators at the μm to mm scale. For some applications (e.g., airfoil surface), the smart skin substrate has to be flexible to conform to the airfoil shape and at the same time has to be compatible with the IC processing for sensor and smart electronics integration. It has been proposed by Carraway[13] that polyimide is an excellent material for use as the skin because of its flexibility and IC processing compatibility. The control loop between the sensors and actuators employs the multifunctional materials that provide electrical functionality at selected locations using conductive polymers and electrodes connected to on-site antennas communicating with a central antenna. A related and difficult problem that has been largely unaddressed is the method for telemetry of the data. In some applications, stresses and strains to which the structure is subjected may pose a problem for conventional cabling. In others, environmental effects may affect system performance. Advances in ultraflat antenna technology coupled with MEMS sensors/actuators seems to be an efficient solution. The integration of micromachining and microelectronics on one chip results in so-called smart sensors. In smart sensors, small sensor signals are amplified, conditioned, and transformed into a standard output format. They may include a microcontroller, digital signal processor, application-specific integrated circuit (ASIC), self-test, self-calibration and bus interface circuits, simplifying their use and making them more accurate and reliable (see Figs. 11.1 and 11.2).

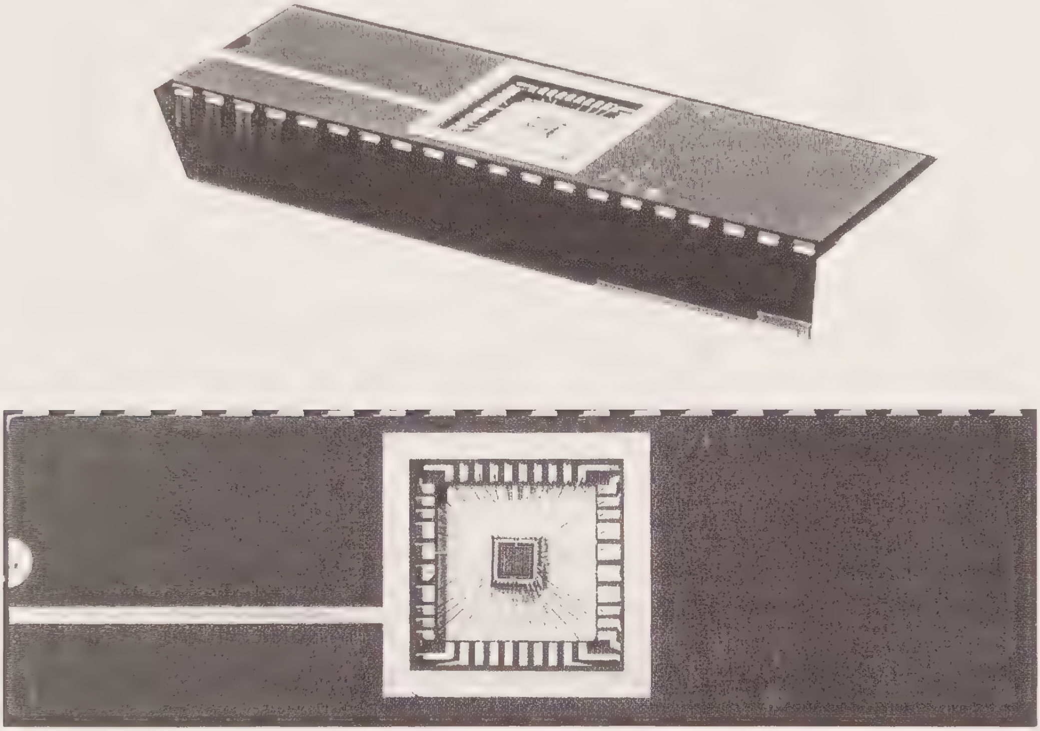


Fig. 11.1. Penn State ASIC chip.

The smart electronics referred to in this paper are microelectronics circuitry such as CMOS, flip-chip electronics, microprocessors, and wireless remote and continuous telemetry systems that integrate various electronic components and microsensors and actuators. The wireless communication obviates the need for complicated wiring to accomplish the telemetry. For such integration the antenna should be compatible with the standard IC technology and micromachined along with other microsensors, microactuators, and control devices. The integrated MEMS actuators with suitable smart electronics etched onto the structure are controlled by the built-in antennas through feedback and feedforward control architecture. For local and global control, all the sensors and actuators in MEMS are independently connected to a central control system by a large amount of wiring. In a realistic system, the star organization is replaced by a bus system, along with ASIC, in which the information is transmitted as a binary signal along a single data bus line. For a centralized control, a telemetry device is usually adopted as shown in Fig. 11.2. In the subsequent integration of multifunctional smart materials, MEMs and smart electronics is presented with some examples pertinent to rotary and fixed wing aircrafts.

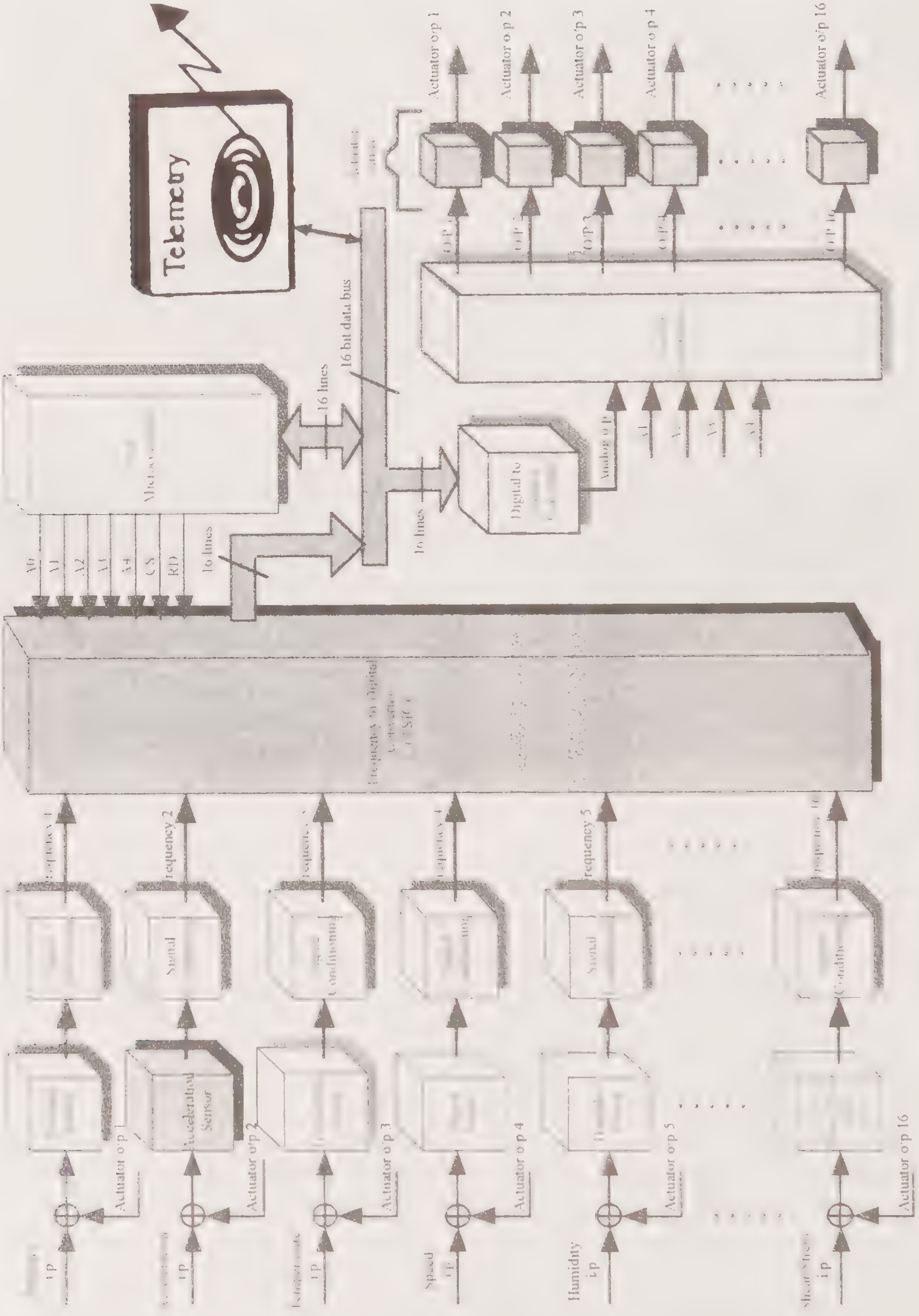

Telemetry
16 bit data bus
16 lines
16 lines
16 lines
A0
A1
A2
A3
A4
CS
RD
Digital to
Analog o/p
O/P 1
O/P 3
O/P 16
Actuator o/p 1
Actuator o/p 2
Actuator o/p 3
Actuator o/p 4
Actuator o/p 16
Frequency to Digital Converter (ASIC)
Frequency 2
Frequency 4
Frequency 5
Frequency 16
Signal
Conditioning
Acceleration Sensor
Actuator o/p 1
Actuator o/p 2
Actuator o/p 3
Actuator o/p 4
Actuator o/p 5
Actuator o/p 16
Speed i/p
Humidity i/p
Shear Stress i/p
i/p

Fig. 11.2. Smart electronics and communication system for smart structure.

11.2 MULTIFUNCTIONAL SMART MATERIALS

Multifunctional polymeric composite materials involve the integration of polymers and nanoceramic particles by chemical bonding as side groups on a polymer backbone. The polymers in general may be chiral with conductive side groups. They consist of a backbone structure with functional groups that serve as anchor points for the metal oxides as well. The chiral groups serve as a passive damper for noise and acoustic fields by converting them into two longitudinal waves (fast and slow) and four components of shear waves (usual shear waves in an elastic continua and left LCP and right RCP circularly polarized waves traveling with two different wave speeds). The nanoparticles such as PZT and PLZT have active surfaces or functional groups that can bond with the polymer chain. The nanoparticles provide the piezoelectric function (and other ferroelectric functions necessary with a family of ferroelectric materials) in the polymer, and the backbone provides mechanical strength and structural integrity, electrical conductivity, etc. Further active sensing and control of noise can be achieved by microsensors and microactuators etched or micromachined on these functional polymers through microprocessors. The conductivity nature of the polymer serves as semiconductor silicon in which MEMS devices can be etched.

A new class of functional polymers are being synthesized at Penn State that can serve as the host for piezoelectric nanoceramics and even Ni/Ti shape memory films, as well as a protective coating layer on structural parts. For these polymers to be used as smart skins or smart devices in structures, they have to meet the following requirements: (1) interactions (chemical or physical) between functional polymer and nanoceramics; (2) strong interfacial adhesion between functional polymer and conducting polymer layers; (3) suitable elastic moduli to support the deformation initiated by MEMS actuator devices; (4) excellent overall dimension stability (allowing local mobility); (5) processes conducive to the attachment of nanoceramics and/or conductive phases and formation of a uniform coating layer; (6) long-term environmental stability.

The functional groups provide strong interactions (chemical, physical, or both) between polymer and piezoelectric nanoceramic or Ni/Ti particles. Some of the functional groups also serve as the anchor sites at the interface of smart composite and conducting polymer layers as well as the active species for crosslinking reaction between polymer chains if needed. The polyolefin polymers, for example, offer good mechanical strength, excellent processability, low moisture permeability, and excellent environmental stability. The elastic property of polyolefin is strongly dependent on the percentage of crystallinity that can be controlled during the polymerization. With the attachment of nanoceramic smart materials, the functional polymer composites form a two-phase morphological structure, a continuous polyolefin elastic phase, and a discrete hard phase that mainly consists of functional groups of piezoelectric nanoceramic particles. The strength of interfacial interactions is governed by the nature and concentration of functional groups.

Many polymers cannot withstand the high electric fields needed to pole ceramics. So it may be preferable to pole the calcined nanoparticles before chemical attachment to the backbone polymer chain. In this way thin films of the nanoparticles held together with a binder can be poled without the need for extremely high voltage poling equipment, which would be the case if thick samples were to be poled. After the thin film samples are poled, the binder can be removed and one now has poled, piezoelectric nanoceramic particles. The surface of the particles must be activated to allow for chemical attachment to the polymer as well as alignment of groups of particles in required directions during attachment.

11.3 MICROELECTROMECHANICAL SYSTEMS

The basic MEMS utilize a diaphragm-based, microbridge-based, or cantilever-based structure. Special processing steps commonly known as micromachining are needed to fabricate these membranes, cantilever beams, and resonant structures, which will be discussed later. For a given application, it may be necessary to have integrated MEMS employing one or more of the basic structures. These three structures provide some feasible designs for the microsensors and actuators that eventually perform the desired task in most smart structures. However, the main issues with respect to implementing these structures are the choice of materials that are to be used in fabricating these devices and the micromachining technology that may be utilized. To address the first issue, we note that in all three of the proposed structures the sensing and actuation occur as a result of exciting a piezoelectric layer by the application of an electric field. This excitation brings about sensing and actuation in the form of expansion in the diaphragm, or in the free-standing beam in the microbridge structure, or in the cantilever beam. In the first two cases the expansion translates into upward curvature in the diaphragm or in the free-standing beam, resulting in a net vertical displacement from the unexcited equilibrium configuration. In the cantilever case, however, and upon the application of an electric field, the actuation occurs by a vertical upward movement of the cantilever tip. Evidently in all three designs the material system structure of the active part (diaphragm, free-standing beam, or cantilever beam) in the microactuator must comprise at least one piezoelectric layer as well as conducting electrodes for the application of the electric field across this layer. Piezoelectric force is used for actuation in many of the applications mentioned above. Micromachining is employed to fabricate the membranes, cantilever beams, and resonant structures.

Microsensors and actuators are fabricated using the well known micromachining techniques in microelectronics industry (for recent developments, please refer to Refs. 1–5. Three-dimensional microactuators in polymer structures can be achieved using stereo lithography on UV curable backbone-type polymers.[6–12] In the integrated MEMS device, we may use photoforming processing in achieving the combined sensor and actuator architecture as outlined by Takagi and Nakajima.[12] For large actuation, one could use a flex tensional transducer consisting of a piezoelectric diaphragm bridged into a cavity.[14]

Silicon micromachining has been a key factor in the vast progress of MEMS during the last decade. This refers to the fashioning of microscopic mechanical parts out of silicon substrates and, more recently, other materials. It is used to fabricate such features as clamped beams, membranes, cantilevers, grooves, orifices, springs, gears, and suspensions. These can be assembled to create a variety of sensors. The Japanese recently exhibited a complete functioning micromachined automobile that operated for several minutes. Bulk micromachining is the commonly used method, but it is being replaced by surface micromachining, which offers the attractive possibility of integrating the machined device with microelectronics that can be patterned and assembled on the same wafer. Thus power-supply circuitry and signal processing using ASICs can be incorporated. It is the efficiency of creating several such complete packages using existing technology that makes this an attractive approach. At present universities have access to several centers that can manufacture to customer design. The Microelectronics Center of North Carolina (MCNC) funded by ARPA, and the NSF-funded Nanofabrication Centers at Cornell University and the Pennsylvania State University are a few.

11.3.1 Micromachining and Microfabrication

Silicon micromachining has been a key factor for the vast progress of MEMS. Silicon micromachining refers to fashioning microscopic mechanical parts out of a silicon substrate or on a silicon substrate. Silicon micromachining comprises two technologies: bulk micromachining, in which structures are etched into a silicon substrate, and surface micromachining, in which the micromechanical layers are formed from layers and films deposited on the surface.

11.3.1.1 Bulk micromachining

Bulk micromachining is the most mature of the two silicon micromachining technologies. It emerged in the early 1960s and has been used since then in the fabrication of different microstructures. It is utilized in the manufacturing of majority of commercial devices: almost all pressure sensors and silicon valves and 90% of silicon accelerometers. The term *bulk micromachining* comes from the fact that this type of micromachining is used to realize micromechanical structures within the bulk of a single-crystal silicon wafer by selectively removing (etching) wafer material. The microstructures fabricated using bulk micromachining may cover the thickness range from submicron to full wafer thickness (200–500 μm) and the lateral size range from submicron to the lateral dimensions of a full wafer. For etching such thick silicon substrates, anisotropic wet etchants such as solutions of potassium hydroxide (KOH), ethylenediamine and pyrocatechol (EDP), tetramethylammonium hydroxide (TMAH), and hydazine-water are used. These etchants have different etch rates in different crystal orientations of the silicon.[2, 15] In most cases wet etching is done from the back side of the wafer while the plasma etching is being applied to the front side. In recent years, a vertical-walled bulk micromachining technique known as SCREAM, which is a combination of anisotropic and isotropic plasma etching, has been used.[16] The etch process can be made selective by the use of dopants (heavily doped regions etch slowly), or may even be halted electrochemically (e.g., etching stops upon encountering a region of different polarity in a biased p-n junction). A region at which wet etching tends to slow down or diminish is called an "etch stop". There are several ways in which an etch-stop region can be created: doping-selective etching (DSE) and bias-dependent DSE.[2, 5, 16]

11.3.1.2 Surface micromachining

Surface micromachining does not shape the bulk silicon, but instead builds structures on the surface of the silicon by depositing thin films of "sacrificial layers" and "structural layers" and by eventually removing the sacrificial layers to release the mechanical structures. The dimensions of these surface micromachined structures can be several orders of magnitude smaller than bulk-micromachined structures. The prime advantage of surface-micromachined structures is their easy integration with IC components, since the wafer is also the working area for IC elements. It should be noted that since miniaturization is immensely increased by surface micromachining, the small mass structure involved may be insufficient for a number of mechanical sensing and actuation applications.

Surface micromachining requires a compatible set of structural materials, sacrificial materials, and chemical etchants. The structural materials must possess the physical and chemical properties suitable for the desired application. In addition, they must have satisfactory mechanical properties; e.g., high yield and fracture stresses, minimal creep and fatigue, and good wear resistance. The sacrificial materials must have good mechanical properties to avoid device failure during fabrication. These properties include good adhesion and low residual stresses in order to eliminate device failure by delamination and/or cracking. The etchants to remove the sacrificial materials must have excellent etch selectivity and must be able to etch off the sacrificial materials without affecting the structural ones. In addition the etchants must have proper viscosity and surface tension characteristics. The common IC compatible materials used in surface micromachining are:

1. Poly-Si/silicon dioxide; LPCVD-deposited poly-Si as the structural material and LPCVD-deposited oxide as the sacrificial material. The oxide is readily dissolved in HF solution without the poly-Si being affected. Together with this material system, silicon nitride is often used for electrical insulation.

2. Polyimide/aluminum; in this case polyimide is the structural material and aluminum is the sacrificial material. Acid-based etchants are used to dissolve the aluminum sacrificial layer.
3. Silicon nitride/poly-Si; silicon nitride is used as the structural material, while poly-Si is the sacrificial material. For this material system, silicon anisotropic etchants such as KOH and EDP are used to dissolve poly-Si.
4. Tungsten/silicon dioxide; CVD-deposited tungsten is used as the structural material with oxide as the sacrificial material. HF solution is used to remove the sacrificial oxide. Other IC-compatible materials such as silicon carbide, diamondlike carbon, zinc oxide, and gold are also used.

Surface micromachining can also be performed using dry-etching methods. Plasma etching of the silicon substrate with SF6/O2-based and CF4/H2-based gas mixtures is advantageous, since high selectivities for photoresist, silicon dioxide, and aluminum masks can be achieved. However, when using plasma etching, a large undercut of the mask is observed. This is due to the isotropic fluorine atom etching of silicon, which is known to be high compared with the vertical etch induced by ion bombardment. In contrast, reactive ion etching of poly-Si using a chlorine/fluorine gas combination produces virtually no undercut and almost vertical etch profiles when using photoresist as a masking material. Thus, rectangular silicon patterns up to 30 mm deep can be formed using chlorine/fluorine plasmas out of poly-Si films and the Si wafer surface.

11.3.1.3 Stereo lithography

Stereo Lithography is a poor man's LIGA for fabricating high aspect ratio MEMS devices in UV-curable semiconducting polymers. Microfabrication of real 3D polymeric and metallic MEMS devices has been accomplished by Ikuta and Hirowatari[10] using stereo lithography on UV-curable polymers. Three-dimensional microstructure of high aspect ratio has been constructed by stacking 2D slices of UV-hardened polymers. The stereo lithography apparatus designed by them is shown in Fig. 11.3.[10] It consists of a UV light source, XYZ positioner and controller, shutter, lens, and microcomputer. The UV-curable polymer is kept in a container. First thin 2D slices are formed by focusing the UV beam with the lens in the XY plane. Fine focusing is necessary to fabricate microsize 2D slices. The focus point is then moved up along the Z axis and the polymer is again cured in that plane. Complicated 3D structures can thus be formed by stacking thin 2D microsize slices (see Fig. 11.4).[10]

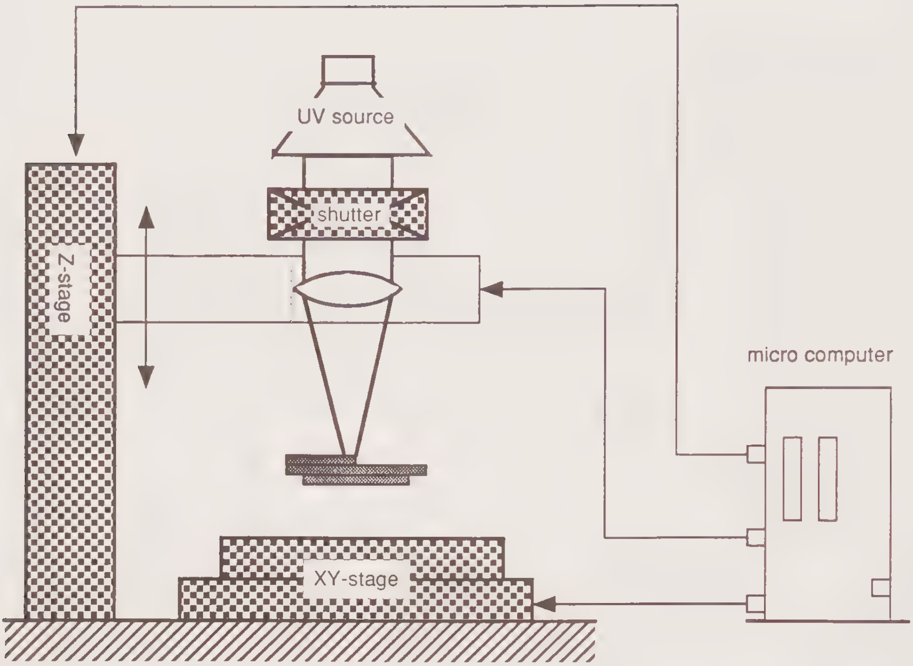

UV source
shutter
Z-stage
micro computer
XY-stage

Fig. 11.3. Stereo lithography unit [10].

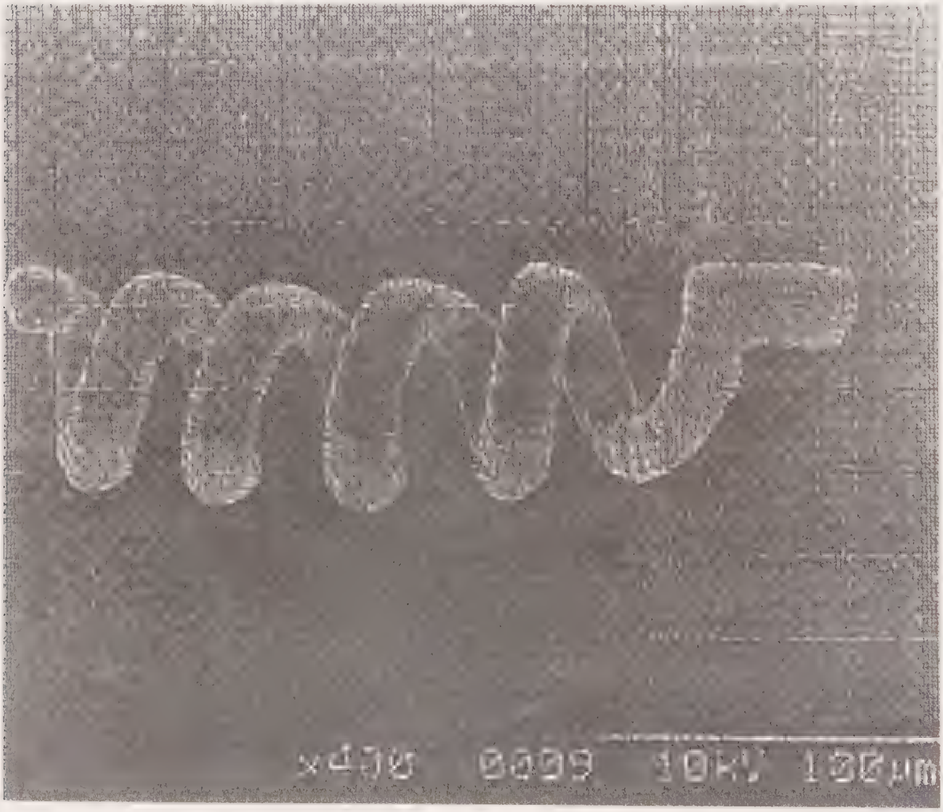

x400 0009 10KV 100μm

Fig. 11.4. Typical polymeric spring [10].

For 3D MEMS devices, the polymers need to have conductive and possibly piezo- or ferroelectric properties. Such MEMS polymers involve the integration of conventional UV-curable polymers, optically transparent conductive polymers, and nanopiezo or ferroelectric particles by chemical bonding as side groups on the polymer backbone. The concept is to design a backbone with functional groups that will serve as anchor points for the metal oxides. The nanoparticles such as PZT and PLZT must have active surfaces or functional groups that can bond with the polymer chain. The nanoparticles provide the piezoelectric function in the polymer and the backbone provides the mechanical stability and flexibility if it is needed.

Stereo lithography can also be used for fabricating composites consisting of microcoiled polymeric and ceramic springs. They have applications in electromagnetic absorption and shielding, tunable acoustic absorbers, drag control, beam focusing and steering, microsensing devices, etc. With suitable conductive and piezoelectric properties incorporated in them, they may serve as excellent MEMS devices. MEMS devices made of these polymers can thus be activated by applying voltage to initiate either a capacitative or piezoelectric effect. Active microcoil springs can thus be fabricated.

In many applications, the precise control of the width and direction of a laser or microwave beam is crucial. Narrowing the beam width results in a power density that is several orders of magnitude higher than the original. A tunable “mirror” is being designed using the multifunctional polymer mentioned above. Stereo lithography is ideal for such fabrication. A tunable doubly periodic rough surface tuned as shown in Fig. 11.5 will reflect a beam in a specific direction with considerable reduction in width.[17]

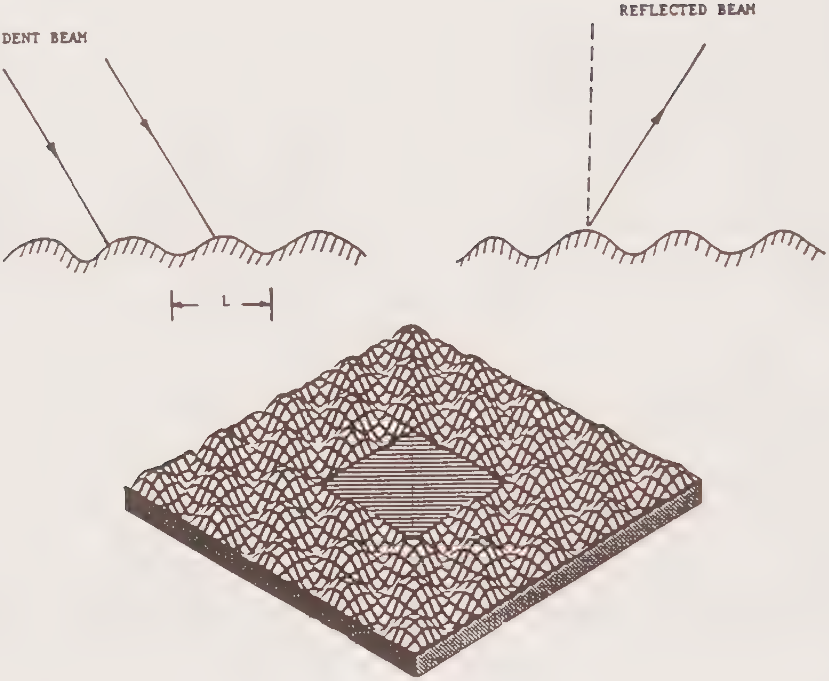


Fig. 11.5. Surface reflectors.

11.4 SMART ELECTRONICS AND CONFORMAL ANTENNAS

The development of integrated smart electronics involves microelectronics circuitry, low-power signal processing electronics, a microcontroller, and wireless communication between microsensors and microactuators. Three different approaches are being pursued: (1) CMOS microelectronics followed by MEMS device integration using polysilicon surface micromachining,[18] (2) micromechanics followed by CMOS integration,[18] (see Figs. 11.6 and 11.7), and (3) flip-chip integration of MEMS chips with that of microelectronics circuitry chips,[19] (see Fig. 11.8). The flip-chip integration is very attractive for both silicon and polymeric MEMS devices. The electronics are fabricated on silicon substrates; the MEMS may be fabricated either on silicon or on polymeric substrates. Then the electronics and MEMS are connected using solder bumping.

A built-in electronics system called BIFET process has been conceived at Delft University of Technology.[20] A simplified version with surface acoustic wave devices is shown in Fig. 11.9.[21] The application of this smart electronics to IDT microsensors and wireless strain and accelerometers is presented later.

Polyimide coatings of 1–2 mm may also be compatible with the smart skin, which will be discussed later. For local intelligence and decision making, one could add ASIC to the electronics as outlined in Figs. 11.1 and 11.2.

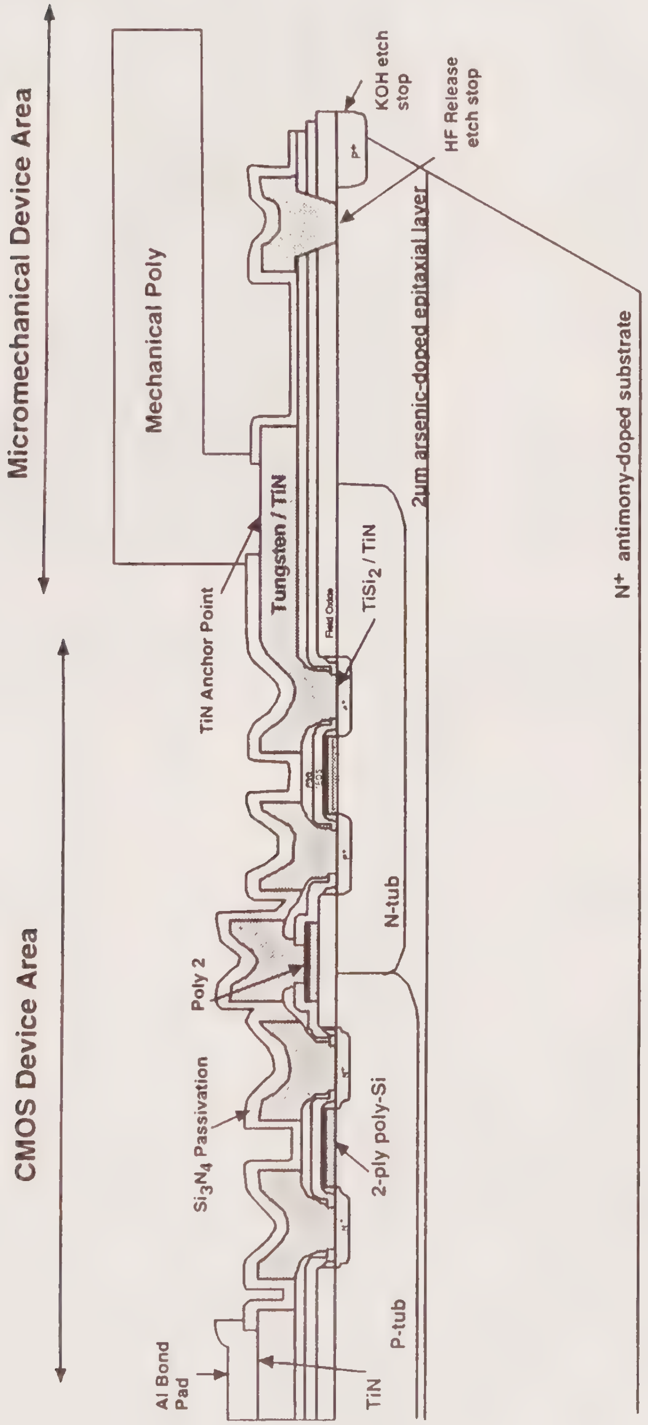

Micromechanical Device Area
CMOS Device Area
Mechanical Poly
KOH etch stop
HF Release etch stop
p+
2um arsenic-doped epitaxial layer
N^+ antimony-doped substrate
Tungsten / TiN
TiN Anchor Point
$TiSi_2$ / TiN
Field Oxide
N-tub
Poly 2
Si_3N_4 Passivation
2-ply poly-Si
P-tub
Al Bond Pad
TiN

Fig. 11.6. CMOS first - micromechanics second [18].

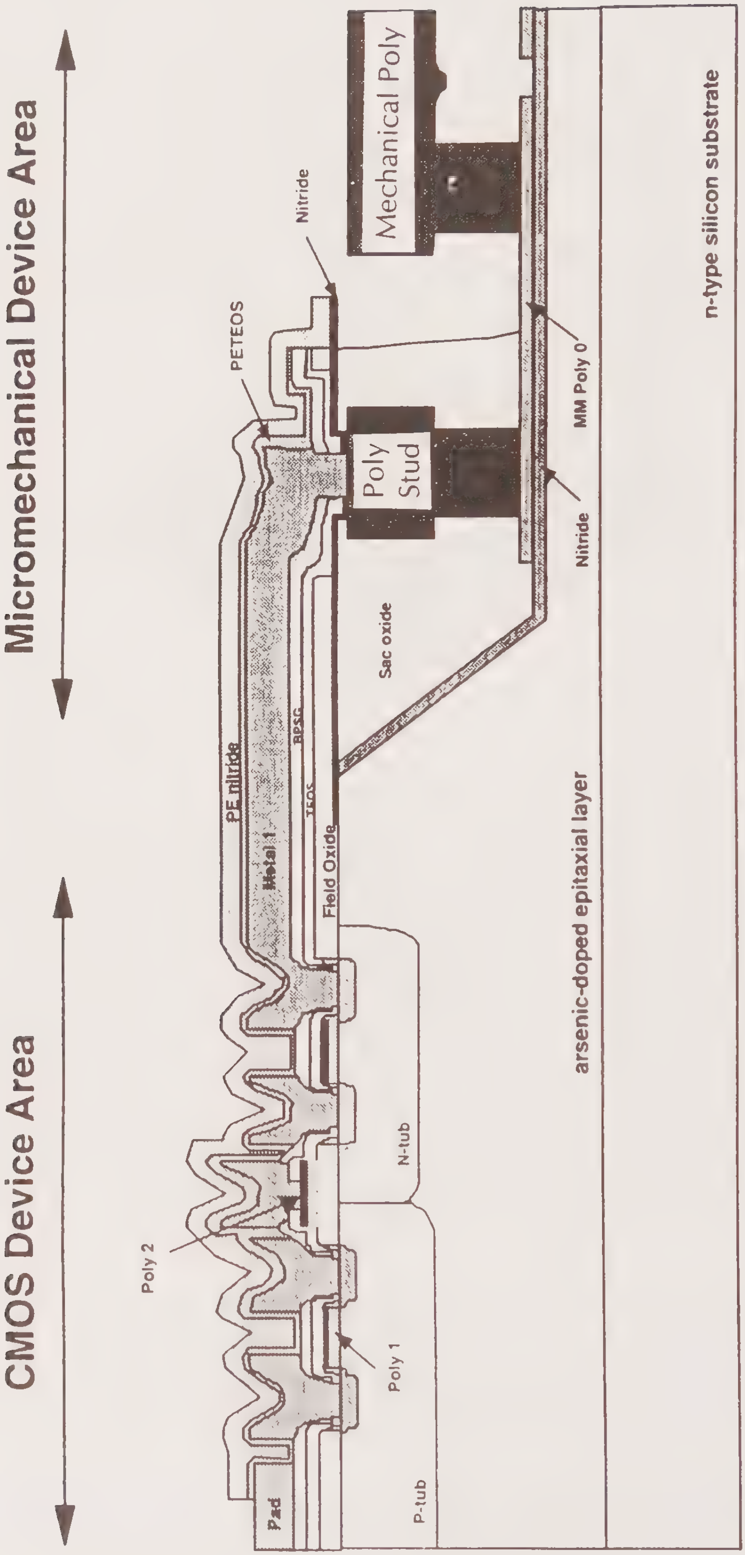

Micromechanical Device Area
CMOS Device Area
Mechanical Poly
Nitride
PETEOS
Poly Stud
MM Poly 0
Nitride
Sac oxide
PE nitride
BPSG
TEOS
Metal 1
Field Oxide
n-type silicon substrate
arsenic-doped epitaxial layer
N-tub
Poly 2
Poly 1
P-tub
Pad

Fig. 11.7. Micromechanics first -CMOS second [18].

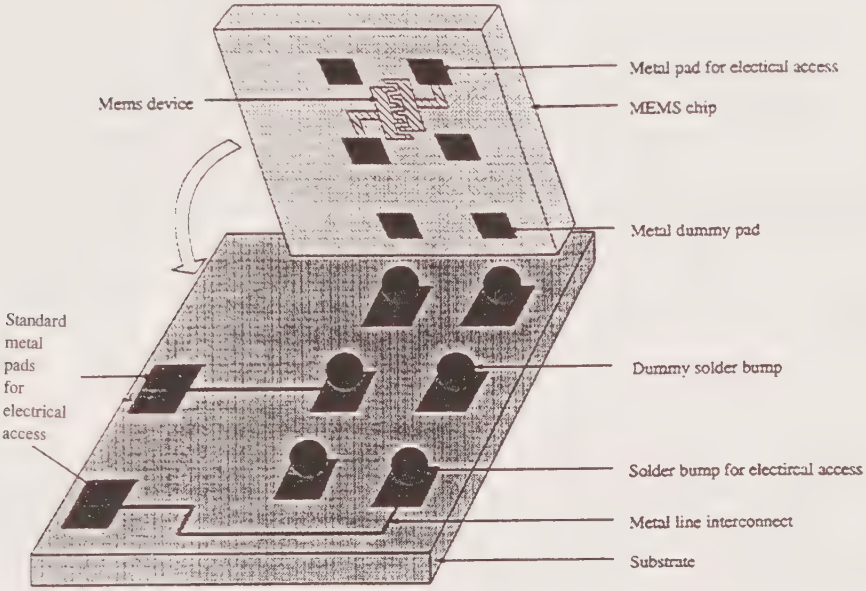

Metal pad for electical access
Mems device
MEMS chip
Metal dummy pad
Standard metal pads for electrical access
Dummy solder bump
Solder bump for electircal access
Metal line interconnect
Substrate

Fig. 11.8. Flip-chip bonding [19].

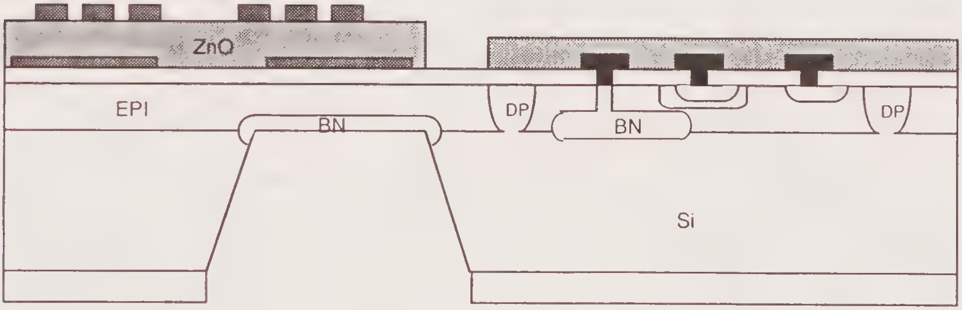

ZnO
EPI
BN
DP
BN
DP
Si

Fig. 11.9. Delft field transistor [21].

11.4.1 Application-Specific Integrated Circuits

An ASIC is a semiconductor device designed for a specific application. The two major categories of ASIC technology are array based and cell based. Array-based ASICs configure a specific design at metal layers, whereas cell-based ASICs are uniquely fabricated at all layers of the silicon process, including the diffusion layers. The chip has a wide range of applications. In conjunction with sensors such as accelerometers, it can be used to design smart sensors. The chip can play a vital role in engine controllers and in pressure measurements using vibrating-type transducers. Although relatively stable frequency standards have been available for many years, precise frequency measurement has not always been an easy task. Early frequency measurement required precision standards, frequency comparators, and interpolation oscillators, as well much operator skill. This came to an abrupt end with the introduction of digital logic and development of the frequency counter. In general, the frequency is measured using two general techniques: direct frequency measurement and period measurement. We have developed a hybrid technique with high accuracy.

The hybrid technique makes use of period and frequency measurement methods to reduce gating error and to keep the clock frequencies below 10 MHz. It is very effective for conversion of frequency into digitized data over a wide range of frequencies from 100 Hz to 100 KHz. This is a novel method not available in general laboratory instrumentation. The block diagram of the hybrid technique is shown in Fig. 11.10.

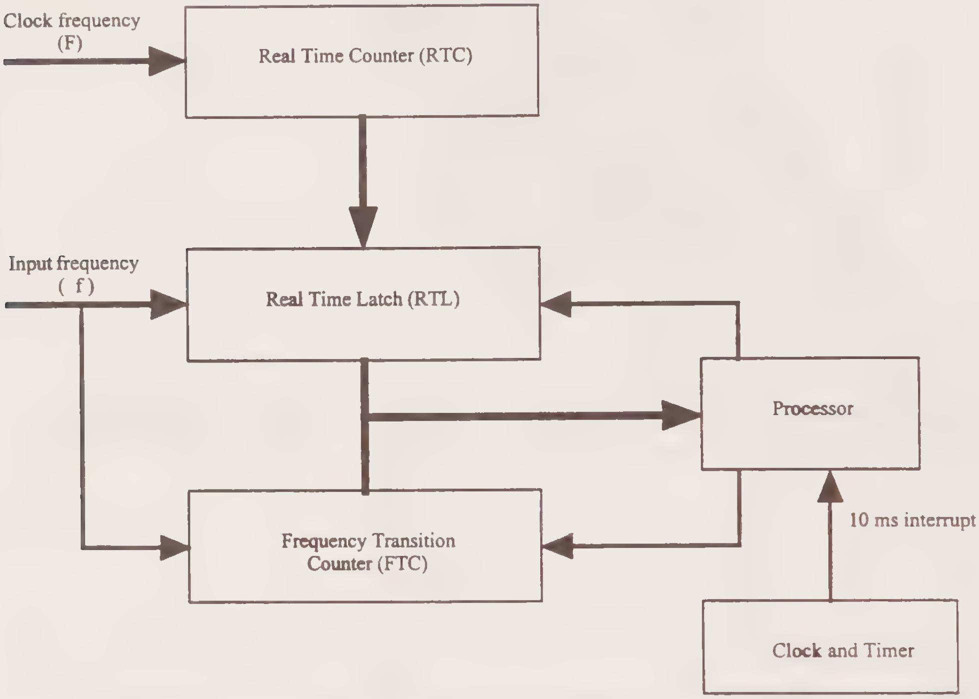

Clock frequency
(F)
Real Time Counter (RTC)
Input frequency
(f)
Real Time Latch (RTL)
Processor
10 ms interrupt
Frequency Transition
Counter (FTC)
Clock and Timer

Fig. 11.10. Block diagram of the hybrid technique of frequency measurement.

In this technique, two counters and a latch are employed. One counter, called the real-time counter (RTC), counts a clock frequency (FMHz). Another counter, called the frequency transition counter (FTC), counts the positive transitions of the unknown input signal *f*. A real-time latch (RTL) latches the RTC value at each positive transition of the input signal. The processor reads RTL and FTC values at every 10-ms interval. The measured values are stored in the processor, and using these values the frequency of the input signal is computed.

Let the values measured at two consecutive periods be RTL_{N-1}, FTC_{N-1}, and RTL_N, FTC_N. Then the number of positive transitions that occurs between N-1 and N read by the processor is:

$$FTC_N - FTC_{N-1} \quad . \tag{11.1}$$

The time interval between the N−1 and the N readings, read by the processor is:

$$\frac{\mathrm{RTL}_{\mathrm{N}}-\mathrm{RTL}_{\mathrm{N}\ 1}}{F} \quad . \tag{11.2}$$

The frequency of the input signal is therefore

$$\frac{\mathrm{FTC}_{\mathrm{N}}-\mathrm{FTC}_{\mathrm{N}-1}}{\mathrm{RTL}_{\mathrm{N}}-\mathrm{RTL}_{\mathrm{N}-1}}\times F \quad . \tag{11.3}$$

The hybrid technique uses the computational capability of the processor to read the FTC and RTL values at 10-ms intervals and computes the frequency. Normally, the processor generates this 10-ms period by programming a counter/timer to generate an interrupt at the required rate. The interrupt latency of the processor depends on the instruction the processor is executing and the priority of other interrupts that may be pending at that moment. The accuracy of this period is normally 10 ms. In the computation of frequency, the processor reading is not a parameter and hence its accuracy does not affect the accuracy of frequency measurement.

11.4.1.1 Accuracy of the hybrid technique

The accuracy of the hybrid technique for a frequency of 100 KHz, for example, is computed as follows:

$$f = \frac{\text{Transitions of the unknown frequency in 10ms}}{\text{Transitions of the clock frequency in 10ms} \pm 1 \text{ count}} \times \text{Clock frequency} \tag{11.4}$$

If the clock frequency used is 2 MHZ this becomes

$$= 100000 \pm 5 \text{ Hz} \tag{11.5}$$

$$\%\ \text{Error} = \frac{5}{100000} \times 100$$

$$= 0.005\% \tag{11.6}$$

The accuracy of measurement remains constant over the entire input frequency range. This is because at higher frequencies the frequency is not measured using one cycle of the input signal; it is averaged over a number of transitions that occur during a 10 ms period.

If the clock frequency F is fed to the RTC and the unknown frequency f is applied to the RTL and the FTC, the RTC counts the number of transitions of the clock frequency and the FTC counts the number of transitions of the unknown frequency. The RTL latches the count value of the clock frequency at every upward transition of the unknown frequency. The waveforms are shown in Fig. 11.11.

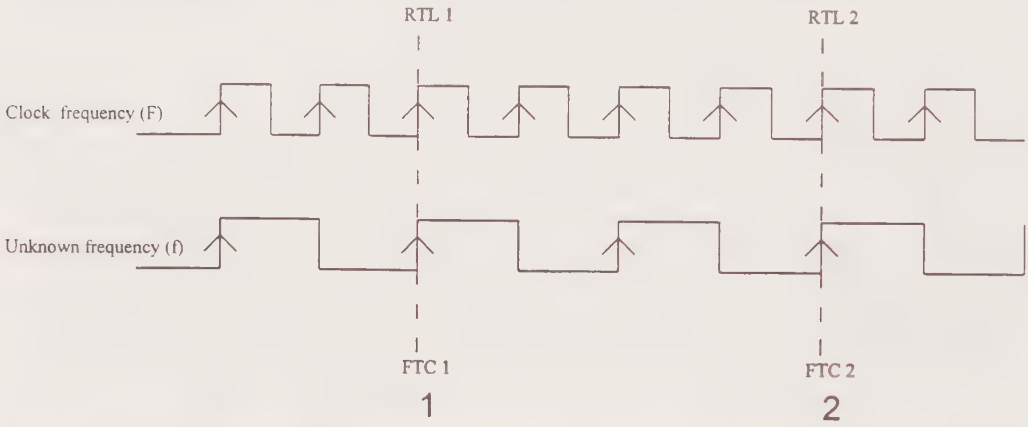


Fig. 11.11. Waveforms showing the method of measurement in the hybrid technique.

At any instant 1 the RTL and FTC values are RTL1 and FTC1. At instant 2, the respective values are RTL2 and FTC2. The frequency f is calculated as given below: Time period of the unknown input signal is

$$t = \left(\frac{\mathrm{RTL2}-\mathrm{RTL1}}{\mathrm{FTC2}-\mathrm{FTC1}}\right) \times \frac{1}{F} \quad . \tag{11.7}$$

Therefore the unknown input frequency is

$$f = \left(\frac{\mathrm{FTC2}-\mathrm{FC1}}{\mathrm{RTL2}-\mathrm{RL1}}\right) \times F \quad . \tag{11.8}$$

The block diagram of a single-channel frequency-to-digital converter is shown in Fig. 11.12. The contents of a single-channel frequency-to-digital converter is shown in Fig. 11.13. The inputs to the chip are described below :

- FREQ : This is the unknown input frequency that is to be measured by the chip.

- BITE : This is the signal which activates the built in test equipment[6] in the chip. If this input is made low, the unknown frequency FREQ is measured by the chip. If it is set high, a known test frequency TEST_F is fed to the chip and measured to verify the correctness of operation.

- TEST_F : This is a known test frequency that can be fed to the chip for testing purposes . This is part of the built in test equipment provided in the chip.

- MRST~ : This is the master reset signal, which is active low . If this signal is made low, all the outputs are reset to zero.

- RD~ : This is the read signal, which is active low . When it is made low, the value in the real time latch or the frequency transition counter is read and appears on the 16 output lines, depending on the value of the A0 address bit.

- CS~ : This is the chip select signal which is active low . Only when CS is made low, the channel is selected and the frequency is measured.

- A0 : This is the least significant address bit . The value of A0 determines whether the value of the real time latch or the frequency transition counter appears on the 16 output lines.

- A <4:1> : These are the four address lines which are fed to the address decoder. The four-bit address value determines the selection of the appropriate channel. This is for expansion of the chip to handle multiple channels of frequencies.

- RTC_CLK : This is the clock frequency that is fed to the real time counter in the chip. The clock to the RTC is made optional between 8MHz, 4MHz, 2MHz and 1MHz depending on the accuracy desired.

- 16 MHZ : This is the 16-MHZ frequency that is fed as the clock input to the frequency divider and the state machine in the chip.

The outputs of the chip are described below :

- 8 MHZ, 4 MHZ, 2 MHZ, and 1 MHZ: These are four frequency outputs of the chip which are the optional time base frequencies from which the clock to the real-time counter is selected.

- D <15:0> : These are the 16 output data lines of the chip . The 16-bit data can be the real-time latch value or the frequency transition counter value, depending on whether the A0 address line is 0 or 1.

The hybrid technique necessitates the use of the following functional units in the chip, which have been shown in Fig. 11.13.

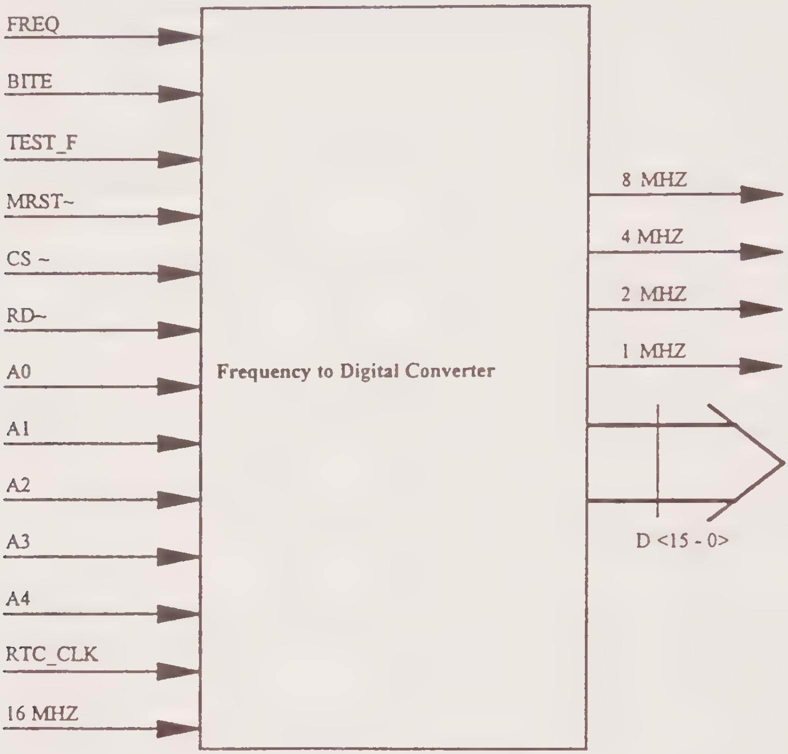

FREQ
BITE
TEST_F
MRST~
CS ~
RD~
A0
A1
A2
A3
A4
RTC_CLK
16 MHZ
Frequency to Digital Converter
8 MHZ
4 MHZ
2 MHZ
1 MHZ
D <15 - 0>

Fig. 11.12. The block diagram of a single channel frequency to digital converter.

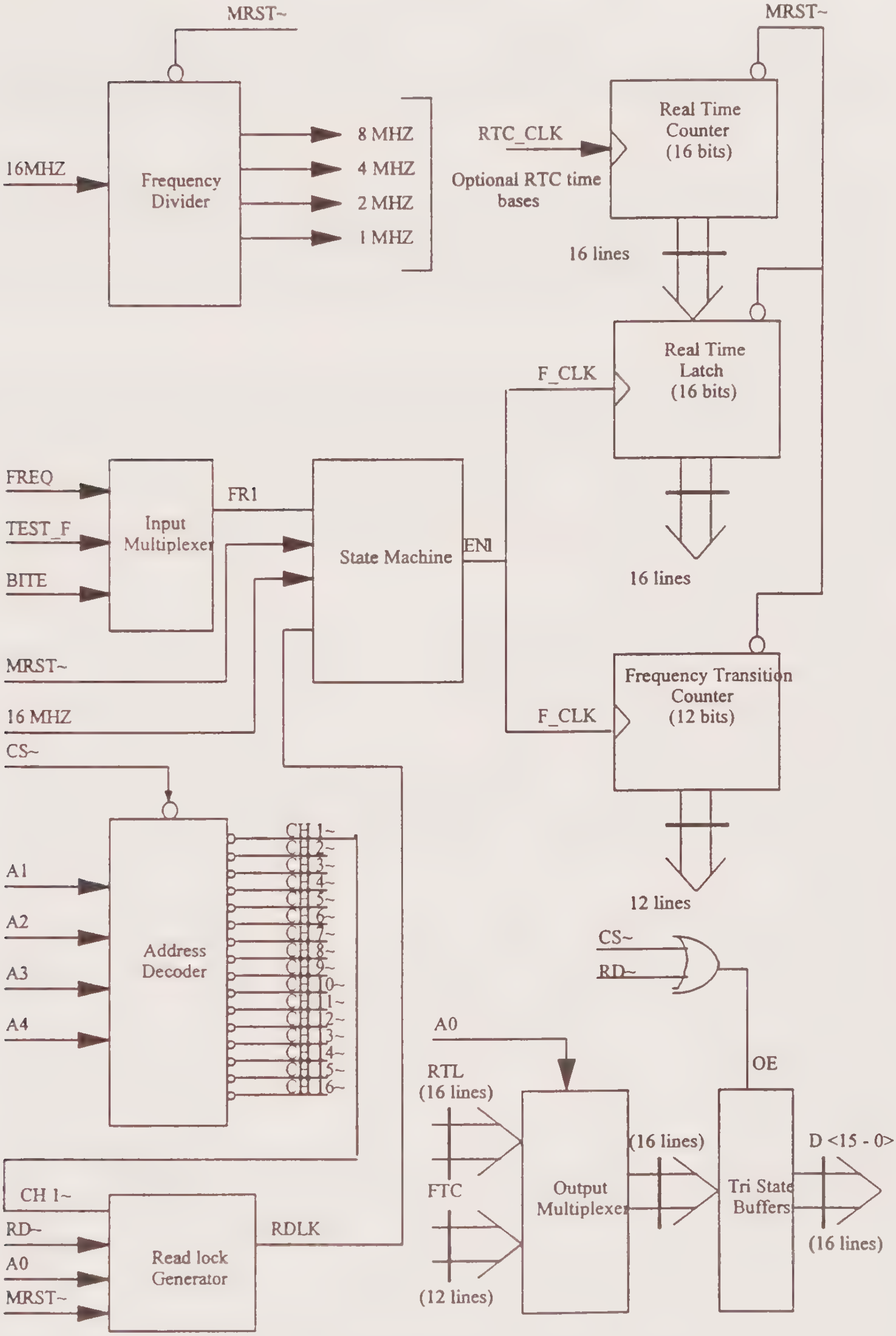

MRST~
MRST~
16MHZ
Frequency Divider
8 MHZ
4 MHZ
2 MHZ
1 MHZ
RTC_CLK
Optional RTC time bases
Real Time Counter (16 bits)
16 lines
Real Time Latch (16 bits)
F_CLK
16 lines
FREQ
TEST_F
BITE
Input Multiplexer
FR1
State Machine
ENl
MRST~
16 MHZ
CS~
Frequency Transition Counter (12 bits)
F_CLK
12 lines
A1
A2
A3
A4
Address Decoder
CH 1~
CH 2~
CH 3~
CH 4~
CH 5~
CH 6~
CH 7~
CH 8~
CH 9~
CH 10~
CH 11~
CH 12~
CH 13~
CH 14~
CH 15~
CH 16~
CS~
RD~
A0
OE
RTL (16 lines)
FTC
(12 lines)
Output Multiplexer
(16 lines)
Tri State Buffers
D <15 - 0>
(16 lines)
CH 1~
RD~
A0
MRST~
Read lock Generator
RDLK

Fig. 11.13. Contents of a single channel frequency to digital converter.

In an address decoder, the 16 active low outputs are the 16 channel selects. This has been provided for future expansion of the chip to handle up to 16 frequency channels. The decoder inputs are the address lines A <4:1> (four address lines).

Depending on the value of the address the appropriate channel gets selected. The decoder also has a chip select that is active low. The decoder operates only when the chip select CS is set low.

The input to the frequency divider is a 16-MHZ 50% duty cycle waveform. This is to be divided to produce four output frequencies of 8 MHZ, 4 MHZ, 2 MHZ, and 1 MHZ, which are all 50% duty cycle waveforms. The output frequencies generated by the frequency divider are used as optional time base frequencies to the 16-bit real-time counter.

The input multiplexer is basically a 2 to 1 multiplexer. This is present at the input to provide a built in test equipment (BITE). There is a BITE <0> input connected to the select line of the multiplexer. If this input is set low, the unknown input frequency appears at the output of the multiplexer. If this line is set high, then it passes the test frequency that verifies the operation of the chip, thus activating the self-test mechanism in the chip.

The frequency transition counter is a 12-bit synchronous counter[24] that counts the cycles of the unknown frequency. The output of the state machine is fed as the clock to the frequency transition counter.

The real-time counter is a 16-bit synchronous counter that counts the cycles of the standard time base frequency. The standard time base frequency can be selected from 8-MHZ, 4-MHZ, 2-MHZ, or 1-MHZ values. These optional time bases are used depending upon the accuracy to be attained and to avoid excessive overflow.

The real-time latch is a 16-bit latch that latches on to the value of the real-time counter at every transition of the unknown frequency input. The state machine`s output is fed as the clock input to the real-time latch.

The read-lock circuit generates an output signal RDLK that remains high over the entire duration of the read cycle. The generation of this signal is very essential because when the microprocessor wants to read the values of the RTL and the FTC, it asserts the read line low and sends the appropriate value of the address bit A0. When the address bit is low, the RTL is read and when it is high, the FTC is read. During this reading time, it is required to prevent the FTC from updating its previous value. This is achieved by generating the RDLK signal.

The function of the state machine is to arbitrate suitably between the frequency output of the input multiplexer FR1 and the read-lock signal generated by the read-lock circuit RDLK and produce an output EN1 that is fed as the clock to the RTL and the FTC.

Arbitration is essential because it is necessary to maintain a one-to-one correspondence between the values of the RTL and the FTC. If this is not so, the computed frequency will be erroneous.

The state machine designed is a synchronous Moore machine.[23] It relies on the clock signals to trigger the state transitions. The state machine is positive-edge triggered with respect to the 16-MHZ clock signal. This is to ensure that the inputs are stabilized before the rising edge of the clock arrives.

The input signal FR1 <0> is fed to a synchroniser[23, 24] and latched with the falling edge of the 16-MHZ clock to prevent the rare occurrence of metastability.[23, 24] The read-lock signal RDLK is also latched with the falling edge of the 16-MHZ clock.

The output of the state machine EN <0> is fed to an output-holding register[22] and latched with the falling edge of the 16-MHZ clock to generate the final output EN1 <0> . This is done to prevent the rare occurrence of false state transitions due to the present state output and the state machine output changing at the same edge.

The output multiplexer is basically a 2 to 1 multiplexer. One of the inputs is from the RTL and contains the 16-bit latch value. The other input is from the FTC and contains the 12-bit counter value. The value of the least significant address bit A0 determines whether the latch value or the counter value appears at the output. The output of the multiplexer has 16 lines. When the 12-bit counter value appears at the output, the four most significant bits in the 16-bit output are made zero.

The tristate buffer is a 16-bit tristate buffer provided for purposes of isolation between the outputs of the output multiplexer and the outputs of the chip. The tristate buffer passes the data from its input lines to its output lines only when chip select CS and RD go low simultaneously; otherwise the outputs are in the high impedance state. The 16-bit output of the tristate buffer is the 16 bit output of the frequency-to-digital converter chip.

The present ASIC thus adopts an innovative and effective frequency measurement technique called the *hybrid method*. This method possesses the merits of both the direct frequency measurement and period measurement techniques, so the method is very effective at both low and high frequencies. The chip has a 16-bit RTC, 16-bit RTL, and a 12-bit FTC. An accuracy much better than 0.1% is observed at all frequencies on increasing the time base to the RTC. However, if the clock frequency to the RTC is increased beyond 8 MHZ, there may be considerable overflow of the real time counter when measuring a frequency of 100 KHz by reading every 10 ms. To avoid this, the number of bits in the real time counter may have to be increased to achieve an accuracy of the order 0.0001%.

Thus the chip design is the best possible implementation satisfying all the above conditions. The accuracy can always be improved by increasing the sizes of all the counters and latches. This, however, will lead to a larger layout.[23] The chip was tested extensively for various values of input frequencies, and the hybrid technique was found to be very effective at all frequencies. The accuracy obtained was 0.001% for all frequencies in the range from 100 Hz to 100 KHz, using a time base of 2MHz to the RTC. The time base and the sampling interval were varied, and frequencies much higher than 100 KHz were measured with the same accuracy. The frequency-to-digital converter was further expanded to handle 16 channels of input frequencies.

The applications for which the ASIC is being tested at Penn State include the measurement of speed, acceleration, temperature, pressure, rotation, rotorcraft flex beam deflection, ice sensing and deicing of aircraft, aircraft engine control, measurement of fuel level in aircraft tanks, and measurement of air data in aircrafts. Figure 11.2 illustrates an integration of ASIC with smart devices handling as many as 16 channel different input signals.

11.4.2 Wireless Telemetry Systems

Wireless remote and continuous telemetry systems for application to smart structures can be achieved using Penn State's novel antenna architecture using ferroelectric on-line or filter phase shifters to interact with MEMS, piezoelectric polymer skins, and SAW sensors; the communication from sensors to control systems and back to the actuators is accomplished by wireless telemetry using antennas specially designed for such purpose and integrated into the structure.

The ferroelectric ceramics are used to develop new, lightweight, flat, large-bandwidth antennas whose output beams can be steered electronically (beam steering) by tuning the dielectric properties of the ceramic. The two-plane electronics scanning of the phased array depends largely on phase shifters. Phased arrays, which are required to scan a wide field of view ($\geq \pm 30^{\circ}$) is very expensive, because conventional phased-array antenna design required phase shifters for each radiating element; consequently if there are *N* radiating elements, *N* phase shifters and *N* phase shift drivers are required.

Standard phased-array design consists of a group of radiating elements along with a feed system containing individual phase shifters controlled by PIN diodes or ferrite phase shifters. Phase shifters are currently advanced to monolithic and hybrid. The most commonly used switching element is the silicon PIN diode. Ferrite phase shifters for the phased arrays are built as single units. Radars are typically operating with hundreds of kilowatts to megawatts of peak power at a duty cycle of around 0.001. The ferrite phase shifters are too slow for the above application. However, the PIN diode phase shifters, which can handle only tens of watts, are more lossy than ferrites. For example, ferrite phase shifters can have loss under one dB at 44 GHz, while diode phase shifters might have one dB loss at 5 GHz, but several dB loss at 44 Ghz.[25]

The use of printed microstrip antennas as radiating elements in phased arrays leads to significant reduction of size, weight, and production cost. Integration of the feeding network with the phase shifters in the same substrate plane forms a light, compact, flat antenna system, particularly for airborne applications. Conventional microwave phase shifters like PIN diodes and ferrites are bulky and consume power. The superconducting phase shifters require an additional cooling system. In Ref. 26 we have demonstrated the performance of a microstrip phase shifter developed on a tunable substrate whose dielectric constant is a function of applied dc bias voltage. The phase shift is obtained by the relative change in dielectric constant of the material due to the change in the dc biasing voltages.

In Ref. 26, it has been shown that a filter phase shifter will be an ideal one for many types of electronically steerable antennas. Figure 11.14 presents the schematic diagram of the filter phase shifter. The middle conductor along with the duroid substrate is removed, and the tunable substrate ($Ba_{0.45}$ $Sr_{0.55}$ TiO_3) of size 4.0 cm × 0.6 cm and 1-mm thickness (based on the design criteria given above) is machined and inserted in the duroid substrate with push fitting. The remaining gap is carefully filled with polyester resin and BST powder. Microstrip circuits are etched on the substrate with the middle line on BST. The biasing circuits are attached in the middle line with proper care in impedance matching. The polymer layer is used to increase the voltage breakdown.

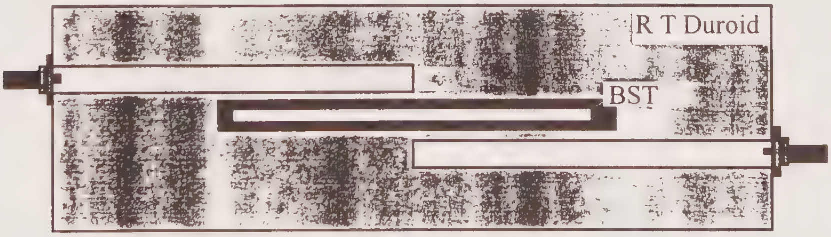


Fig. 11.14. Schematic diagram of the filter phase shifter.

As shown in Fig. 11.15, the filter phase shifter provides higher phase shift when compared with the microstrip phase shifter using DC blocks and chokes, etc.[26] The configuration is easily adaptable with the microelectronics processing. It is evident from the above figures that the performance of the filter phase shifter is much better (20° per kV/cm) compared with the performance of the microstrip line phase shifter reported in Ref. 2 (35° per 12 kV/cm).

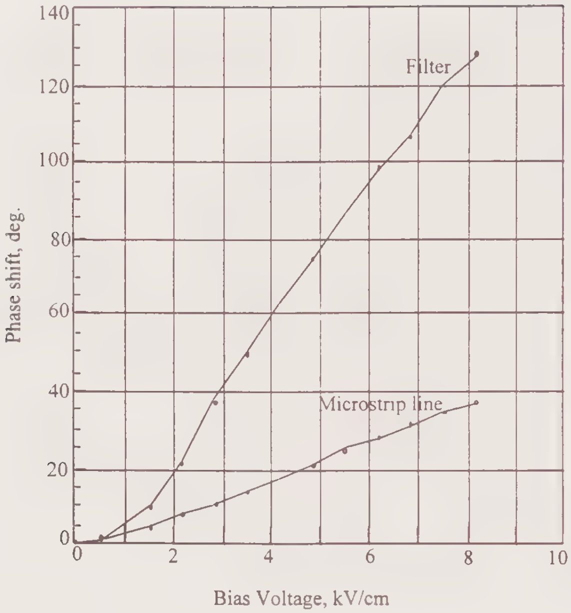


Fig. 11.15. Variation of phase due to the change in bias voltage for the filter and the microstrip line phase shifter.

Thus the ferroelectric films can be deposited on the substrate on which the feed line conductors can be screen printed; an electronically steerable antenna may thus be conceived for beam steering (see Fig. 11.16). Beam steering can be used to advantage for monitoring and communicating with several sensors and actuators distributed over a large structure using a single central antenna (see Fig. 11.17).

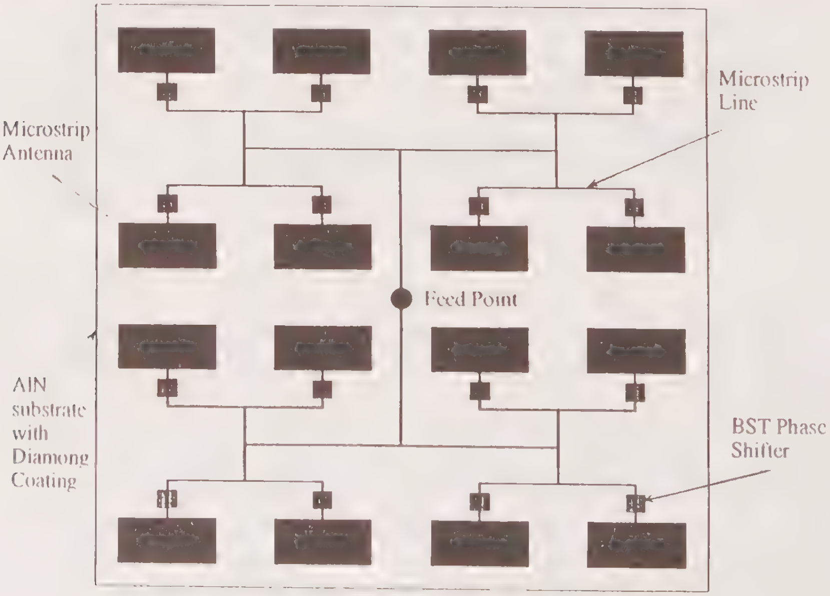

Microstrip Line
Microstrip Antenna
Feed Point
AlN substrate with Diamong Coating
BST Phase Shifter

Fig. 11.16. Microstrip antenna with ferroelectric phase shifters.

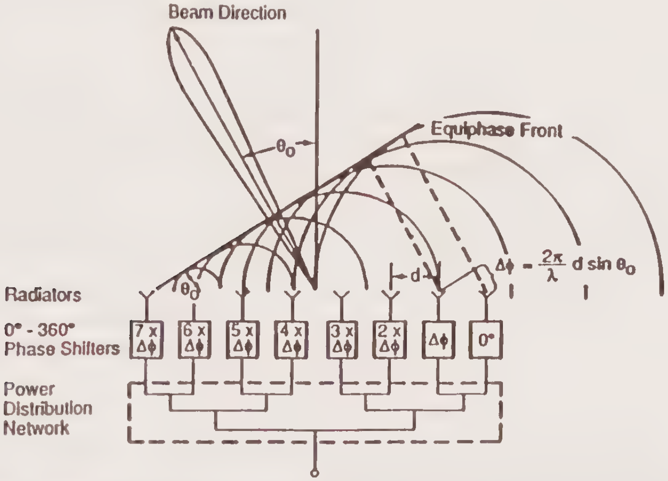


Fig. 11.17. Beam steering concept using phase shifters at each radiating element.

A sensor bus can be used when sensing many parameters at the same time. The sensors can be electronically scanned, and the information is "passed" on to the microcontroller. Separate software code has to be assigned in the microcontroller to each sensor. The wireless communication link consists of a transmitter and receiver with a message format consisting of eight data bits preceded by a starter bit and followed by a parity bit and a stop bit. The antenna for this application is a conformal one employing high-dielectric and low-loss ferroelectric material with tunable capability. Planar phase shifters controlled only by bias voltage are used for electronic steerability.[27–32]

11.5 MICROSENSORS FOR SMART STRUCTURES

Microsensors are basically silicon or polymer devices that convert a mechanical signal into an electronic one by using microelectronics technology. In smart structures, the electronic signal obtained from the sensors is amplified, signal-conditioned, and fed to ASIC chips. Using the intelligence in the electronics, the signal will be processed by the microprocessor and controller and then fed to the actuator. Using this type of integrated or smart sensor, one could achieve many actuation functions locally or remotely, using the telemetry device as shown in Fig. 11.2.

In smart structures, for example in aircraft, one is interested in studying the strain, stress, deflection, pressure, and shear loads on the wings (or rotor blades in the case of helicopters), acceleration, drag and control of drag to reduce the fuel consumption, and associated vibration and noise produced due to the separation of laminar flow to turbulent one, temperature, ice formation and deicing, etc. The measurement of such mechanical signals are also important in the automotive industry, in which the sensors are used in airbags, smart suspensions, control of the correct air-to-fuel ratio in engines, etc.

11.5.1 Interdigital Transducer Microsensors

Recently, micro interdigital transducers (IDT) and micro surface acoustic wave (SAW) sensors are being implemented in structures.[33] They were first used in radar and communication equipment as filters and delay lines, then were found to be attractive sensors for various physical variables such as temperature, pressure, force, electric field, magnetic field, and chemical compounds. A SAW device usually is a piezoelectric wafer with interdigital transducers and reflectors on its surface. The principle of SAW sensors is based on the fact that SAW traveling time between the components, e.g. IDTs and reflectors, changes with the variation of physical variables.

Surface acoustic waves can be generated at the free surface of an elastic solid. The generation of such waves can be achieved by the application of a voltage to a metal film IDT placed on the surface of a piezoelectric solid. The basic configuration of a two port SAW device consists of two IDTs on a piezoelectric substrate. One of these IDTs acts as the device input and converts signal voltage variations into mechanical surface acoustic waves. The other IDT is employed as an output receiver to convert the mechanical SAW vibrations back into output voltages. These devices are reciprocal in nature, as a result signal voltages can be applied to either IDT with the same end result. The outline of a basic SAW device is shown in Fig. 11.18.

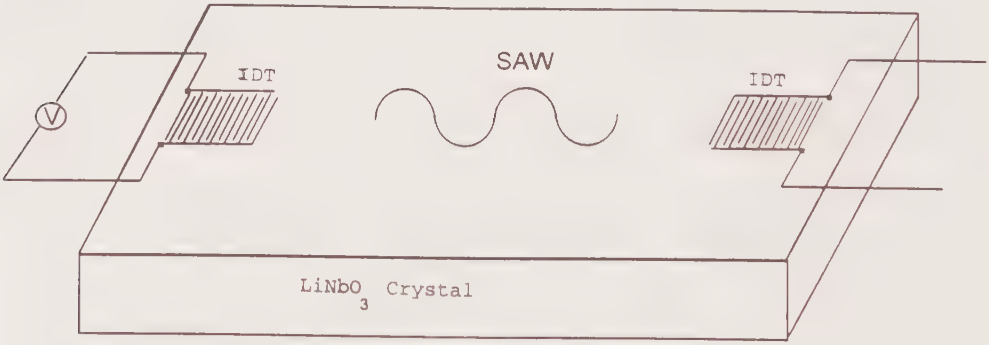


Fig. 11.18. Concept of a SAW device.

A set of IDTs are sputtered or etched on a lithium niobate substrate. They will receive an RF pulse from an antenna to which they are connected. The IDTs will set up the surface acoustic wave on the substrate as a result of this pulse.

For waves propagating in a piezoelectric medium there are two sets of equations, namely the mechanical equations of motion and the Maxwell equations for electrical behavior.

$$\rho\frac{\partial^2 u_i}{\partial t^2} = \sum_{j=1}^{3} \frac{\partial T_{ij}}{\partial x_i} \quad . \tag{11.9}$$

These equations are intercoupled by the constitutive relation listed below.

$$T_{ij} = \sum_k \sum_l C^E_{ijkl} s_{kl} - \sum_k e_{kij} E_k \quad , \tag{11.10}$$

where the superscript on C^E_{ijkl} identifies this as the stiffness tensor for constant electric field; that is, if E is held constant this tensor relates changes of T_{ij} to changes of s_{kl} . The electric displacement D is determined by the field E and the permittivity tensor e_{ij} In a piezoelectric material the electric displacement is also related to the strain. This equation is given below:

$$D_i = \sum_j \epsilon^s_{ij} E_j + \sum_j \sum_k e_{kij} S_{jk} \quad , \tag{11.11}$$

where ϵ^s_{ij} is the permittivity tensor for constant strain.

The constitutive equations for piezoelectric material relating the stress T, strain S, electric field E, and electric displacement D are given by Eqs. (11.10) and (11.11). It can be seen that the electric field and the electric field displacement are coupled in these sets of equations. For a nonpiezoelectric material $e_{kij} = 0$ and there is no coupling between the elastic and the electric fields. The symmetric strain tensor is given by Eq. (11.12), where u is the wave displacement.

$$S_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) \quad . \tag{11.12}$$

The electromagnetic quasi-static approximation is used. The rotational part of the electric field due to the existence of a moving magnetic field is neglected. This approximation [Eq. (11.13)] is valid because the acoustic velocity is small compared to that of the electromagnetic wave. The electric potential is

$$E_i = \frac{\partial \phi}{\partial x_i} \quad . \tag{11.13}$$

with

$$\nabla . D = 0 \quad . \tag{11.14}$$

Using Eqs. (11.11), (11.12), the equations of motion (11.9) in Eq. (11.14) results in

$$\rho \frac{\partial^2 u_i}{\partial t^2} = \sum_j \sum_k \left(e_{kij} \frac{\partial^2 \phi}{\partial x_j \partial x_k} + \sum_l c^E_{ijkl} \frac{\partial^2 u_k}{\partial x_j \partial x_l} \right) . \tag{11.15}$$

$$\sum_i \sum_j \left(\epsilon^s_{ij} \frac{\partial^2 \phi}{\partial x_i \partial x_j} - \sum_k e_{ijk} \frac{\partial^2 u_j}{\partial x_i \partial x_k} \right) = 0 . \tag{11.16}$$

a system of four coupled equations for the electric potential and three components of displacement in a piezoelectric crystal. These are in Eqs. (11.15) and (11.16) where *I, j, k* go from 1 to 3.

The problem of wave propagation on anisotropic substrates can be solved by the method of partial waves. Plane waves solution of the form in Eqs. (11.17) and (11.18) are considered where $j = 1,2,3$ and $m = 1,2,3,4$.

$$u_j^m = \alpha_j^m e^{ikb^m x_3} e^{ik(x_1 - vt)} . \tag{11.17}$$

$$\phi^m = \alpha_4^m e^{ikb^m x_3} e^{ik(x_1 - vt)} . \tag{11.18}$$

The propagation direction is along the x_1 direction. The surface waves decay along the x_3 direction. The wave number is k, b is the decay factor, and v is the phase velocity. The partial wave solutions are substituted into Eqs. (11.15) and (11.16). The weighing coefficients are of these plane waves are chosen to satisfy mechanical and electrical boundary conditions at the surface of the crystal.

The partial wave solutions in Eqs. (11.17) and (11.18) are substituted into Eqs. (11.15) and (11.16) to obtain Eq. (11.19).

$$\begin{bmatrix} m_{11} - \rho v^2 & m_{12} & m_{13} & m_{14} \\ m_{12} & m_{22} - \rho v^2 & m_{23} & m_{24} \\ m_{13} & m_{23} & m_{33} - \rho v^2 & m_{34} \\ m_{14} & m_{24} & m_{34} & m_{44} - \rho v^2 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix} = 0 \tag{11.19}$$

where

$m_{11} = c_{55}b^2+2c_{15}b+c_{11}$
$m_{12} = c_{45}b^2+(c_{14}+c_{56})b+c_{16}$
$m_{13} = c_{35}b^2+(c_{13}+c_{55})b+c_{16}$
$m_{14} = e_{35}b^2+(e_{15}+e_{31})b+e_{11}$
$m_{22} = c_{44}b^2+2c_{46}b+c_{66}$
$m_{23} = c_{34}b^2+(c_{36}+c_{45})b+c_{56}$
$m_{24} = e_{34}b^2+(e_{14}+e_{36})b+e_{16}$
$m_{33} = c_{33}b_2+2c_{35}b+c_{55}$
$m_{34} = e_{33}b_2+(e_{13}+e_{35})b+e_{15}$
$m_{44} = -(\epsilon_{33}b^2+2\epsilon_{13}b+\epsilon_{11})$.

For YZ lithium niobate m_{12}, m_{23} and $m_{24}= 0$. For nontrivial solutions the determinant of the coefficients of a must vanish. For a given value of phase velocity, setting this determinant equal to zero results in an eighth-order equation in the decay constant b. These roots of b are purely real or conjugate pairs. Only values with negative imaginary parts are admissible as these roots lead to waves that decay with depth (i.e., surface waves). There exist four such roots of b. For each of the four roots the corresponding eigenvalues and eigenvectors are determined. A linear combination of the partial waves is then formed as shown in Eq. (11.20).

$$\begin{aligned} U_j &= \left[\sum_m C_m \alpha_j^m e^{ikb^m x_3}\right] e^{ik(x_1 - vt)} \\ \phi &= \left[\sum_m C_m \alpha_4^m e^{ikb^m x_3}\right] e^{ik(x_1 - vt)} \quad , \end{aligned} \tag{11.20}$$

which are then substituted into the boundary conditions. The boundary conditions depend on the given application.

To obtain a high sensitivity, SAW sensors are usually constructed as electric oscillators using the SAW device as a frequency control component. By accurately measuring the oscillation frequency, a small change of the physical variables can be detected by the sensors. A typical SAW oscillator sensor schematic is shown in Fig. 11.19. An amplifier connects two IDTs on a piezoelectric wafer so that oscillations result because of the feedback of the SAW propagating from one IDT to the other. The oscillation frequency satisfies the condition that the total phase shift of the loop equals 2np and varies with the SAW velocity or the distance between the IDTs. The oscillator includes an amplifier and requires an electrical power supply; it cannot be wireless. The operating frequency range of SAW devices is from ten megahertz to a few gigahertz, which directly matches the frequency range of radios and radar. When an IDT is directly connected to an antenna, a SAW device can be excited remotely by electromagnetic waves. The fact suggests that it is possible to realize passive, wireless, or remotely operable SAW devices. The application to remote sensors was first reported in Ref. 34. The temperature of a passive SAW device with a small antenna was remotely read out by a microwave system.[35] In this chapter, a passive, wireless SAW deflection and strain sensor system is described and experimentally tested. The sensing of ice on structures and deicing are also discussed. The analysis is confirmed by experimental results, using a simple model. Additional applications of sensing drag on structures are presented. At Penn State, they have been used for health monitoring of cracks and flaws at or in the vicinity of rivet holes on aircraft surfaces.

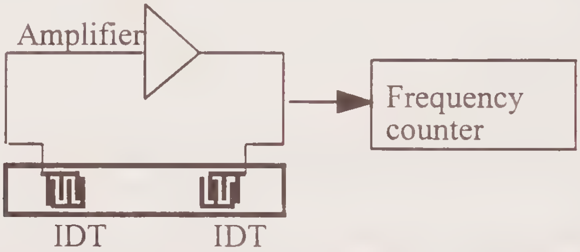

Amplifier
Frequency counter
IDT
IDT

Fig. 11.19. Schematic diagram of an oscillator SAW sensor with a SAW resonator.

The schematic diagram of IDT and reflectors in Fig. 11.20 shows the basic operating principle of the system for wireless communication. An IDT and two reflectors are made on the surface of a piezoelectric crystal wafer. These micro IDT and SAW sensors can be fabricated using a microlithographic process.

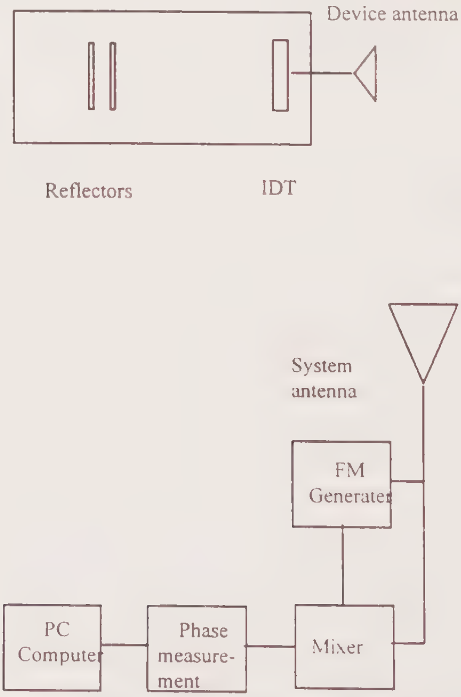

Device antenna
Reflectors
IDT
System antenna
FM Generater
PC Computer
Phase measure-ment
Mixer

Fig. 11.20. Schematic diagram of remote reading sensor system with passive SAW sensor.

The IDT connects directly to a small antenna called the device antenna. This antenna-IDT configuration is able to convert the microwave signal from air to SAW signal on the wafer surface and vise versa. The reading system has a linear frequency modulated (FM) signal generator. The FM signals are transmitted by a system antenna. The signals are received by the device antenna and converted by the antenna-IDT to SAWs propagating along the surface of the wafer. The echoes from the two reflectors are picked up by the antenna-IDT and sent back to system antenna. The echo signals are delayed copies of the transmitted FM signal. The delay times mainly depend on the velocity of the SAWs and distance between the IDT and the reflectors. A mixer, which takes the transmitted FM as reference signal, outputs the signals of frequency difference between the echoes and the transmitted signals. Because the transmitted signal is linear frequency modulated, the frequency difference is proportional to the delay time. By using a spectrum analysis technique like FFT, the two echo signals can be separated in the frequency domain, since the delay times are different. Although the frequencies of the echoes included the information of delay times and the latter are sensitive to the strain on the wafer surface, the sensitivity of frequency to strain is still too low. The phases of the two echoes are also dependent on the delay times and much more sensitive than the frequency, as shown in the following analysis. The system detects the phase difference of the two echoes to figure out the strain, so that the interferences to the common path of the two echoes cancel each other and will not affect the strain measurement.

11.5.2 Microaccelerometer

The microaccelerometer presented in this section was fabricated using surface micromachining. It consists of a seismic mass to measure the acceleration. A 3D picture of the accelerometer is shown in Fig. 11.21. It consists of interdigital finger structures forming a differential capacitor with movable fingers mounted on a movable mass, while the fixed electrodes are anchored onto the silicon substrate. A force-feedback operation mode increases the sensor performance.

In the fabrication process a BiCMOS two-chip approach is adopted. An n+ implantation is performed on a (100) silicon substrate. The purpose of this implantation is to provide an electrical contact to the polysilicon layer that will be deposited on the top of a low temperature oxide layer that acts like a sacrificial layer. A close-loop sensitivity of 8 mV/g and bandwidth of 10 kHz can be achieved.

The active component of many such MEMS devices may involve a moving part that changes the capacitance by changing the spacing between two closely spaced electrodes. The accompanying example of “smart plate” illustrate a MEMS accelerometer that can also be used as a vibration sensor based on changes in electrical capacitance between moving fingers. The design possibilities are almost limitless. The design of a MEMS accelerometer was modified for use as a vibration sensor and was successfully demonstrated as being superior to conventional piezoelectric wafer sensors in sensitivity, occupying far less space and at a fraction of the weight.[36]

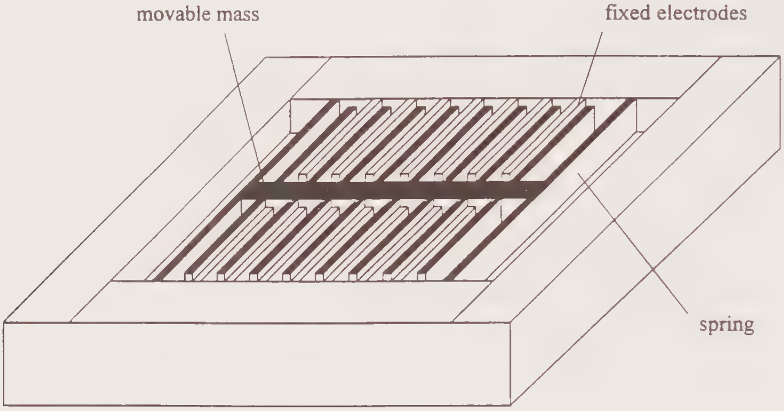


Fig. 11.21. 3-D view of a MEMS accelerometer.

11.5.3 Programmable accelerometer

A programmable accelerometer is very attractive in helicopter applications to study the tip acceleration of rotor blades. It combines the IDT microsensor concept described above and a floating seismic mass on the top of the SAW device as shown in Fig. 11.22. It is amenable to wireless communication of acceleration data from the tip of rotor blades to the antenna in the fuselage of the helicopter. A novel wireless programmable accelerometer has been achieved at Penn State.[37]

The integration of MEMS, SAW devices, and required microelectronics is presented in this section. This unique combination of technologies results in a novel sensor that can be remotely sensed by a microwave system with the advantage of no power requirements or very low power requirements. Such a device is readily compatible with existing antenna technologies, as the SAW device operates at 1 GHz. The microaccelerometer presented is simple in construction and easy to manufacture with existing silicon micromachining techniques. Depending on the application, certain design parameters can be modified to achieve the desired sensitivity .

The relatively small size of the sensor makes it an ideal conformal sensor The accelerometer finds application, in addition to the helicopter, in air bag deployment sensors, vibration sensors for noise control, deflection and strain sensors, inertial and dimensional positioning systems, ABS/traction control, smart suspension, active roll stabilization and four-wheel steering.

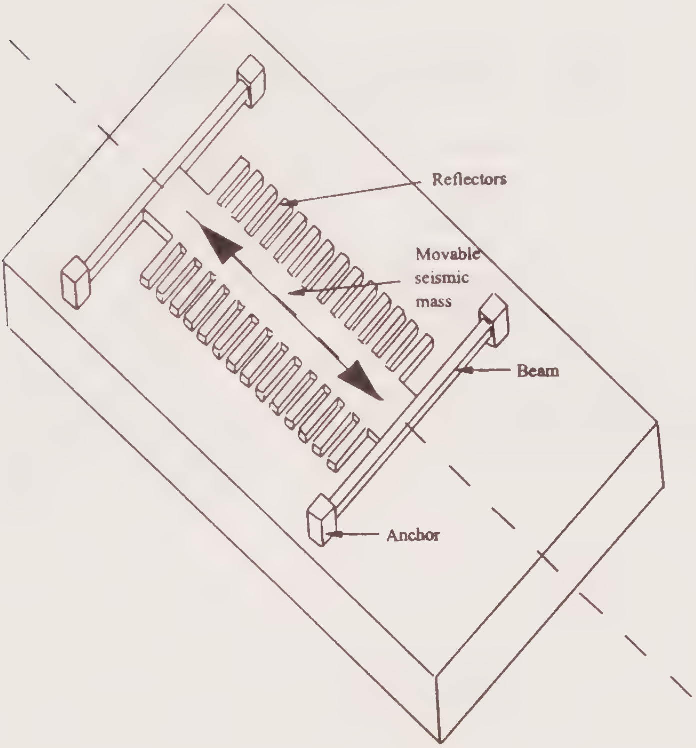


Fig. 11.22. Programmable wireless accelerometer.

Programmable tapped delay lines have used the principle of air gap coupling between the SAW substrate and a silicon superstrate to form individual MOS capacitors. These capacitors are then used to control the amount of RF coupling from the input IDT on the SAW substrate to the output terminal on the silicon chip.[38] This principle has also been successfully implemented in the realization of SAW convolvers.[39] A lithium niobate/silicon air gap coupled MEMS accelerometer is designed using the above principle. The amplitude of the electric potential that the wave induces on the surface of a piezoelectric material extends above the surface by a distance of one acoustic wavelength.

The seismic mass is flip-chip bonded onto the lithium niobate substrate such that the seismic mass is above the substrate by a few tenths of a wavelength. This will ensure sufficient interaction between the electric potential and the reflectors on the seismic mass.

The motion of the seismic mass and reflectors on the mass in response to an acceleration causes a change in this electric potential. There is also a phase shift in the reflected wave due to the new position of the reflectors. This change is now transmitted via the IDT and the antenna. Thus after calibration the acceleration can be read remotely.

A seismic mass similar to that used in silicon accelerometers is used.[40] These are based on the movement of a micromachined seismic mass suspended by a number of narrow beams . A force acting on the mass, due to acceleration, will cause it to move up or down, which causes the suspended beams to strain. This results in an opposing force that increases with increasing mass deflection. Therefore, the position of the mass depends on the momentary acceleration sensed by the device. Such seismic mass structures are frequently used in various designs of accelerometers and have been successfully implemented. Detailed analyses of the dynamic behavior of the seismic mass already exist.[40, 41] For a given geometry of the beam and the size of the mass, the resonance frequency and sensitivity are already determined. Damping of the seismic mass occurs through squeeze film damping. Details of the seismic mass are shown in Fig. 11.23. The structure is realized in polysilicon on a silicon substrate using conventional silicon micromachining techniques.

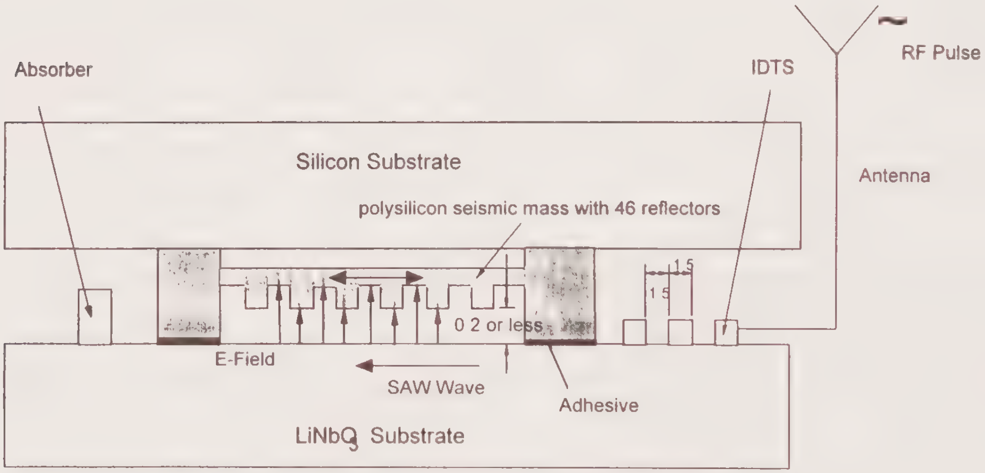


Fig. 11.23. Design configuration of programmable accelerometer.

The design of such an accelerometer follows the Eqs. (11.18)–(11.20) with the appropriate boundary conditions. The applicable boundary conditions for the problem defined are as follows:

1. Mechanical Boundary condition: The surface of the crystal is mechanically free. There is no component of force in the x_3 direction on the surface ($x_3 = 0$). This implies that T_{31}, T_{32}, $T_{33} = 0$.
2. Electrical boundary condition. Since a conductive plate is placed at a height h above the substrate, the potential goes to zero at $x_3 = h$. The potential above the surface satisfies Laplace's equation. The potential and the electric displacement D in the direction normal to the substrate (x_3) are continuous at $x_3 = 0$. This boundary condition is shown in Fig. 11.23 and is represented by equations.

The potential in the air gap is given by

$$\phi_1(x_3) = [Be^{kx_3} + Ce^{-kx_3}]e^{ik(x_1 - vt)} \quad . \tag{11.21}$$

The potential at $x_3 = h$ is zero, resulting in

$$\begin{aligned} \phi_1(h) &= [Be^{kh} + Ce^{-kh}]e^{ik(x_1 - vt)} = 0 \\ C &= -Be^{2kh} \quad . \end{aligned} \tag{11.22}$$

The potential given by Eq. (11.21) is equal to the potential given by the plane wave solution at the surface ($x_3 = 0$). Equating these results in an expression for the unknown constant B:

$$\begin{aligned} \phi_1(0) &= B[1 - e^{2kh}]e^{ik(x_1 - vt)} = \phi(0) \\ \phi(0) &= [\sum_m C_m \alpha_4^m]e^{ik(x_1 - vt)} \\ B &= \frac{\sum_m C_m \alpha_4^m}{1 - e^{2kh}} \quad . \end{aligned} \tag{11.23}$$

The electric displacement in the air gap is given by

$$D_3 = -\epsilon_0 \frac{\partial \phi_1(x_3)}{\partial x_3} \quad . \tag{11.24}$$

The electric displacement on the surface of the crystal is given by

$$D_3(x_3 = 0) = -k\epsilon_0 \frac{\sum C_m \alpha_4^m}{1 - e^{2kh}}[1 + e^{2kh}]e^{ik(x_1 - vt)} \quad . \tag{11.25}$$

This equation is obtained from the potential equation (11.20). The expression for the electric field displacement from the plane wave solution of potential is similarly obtained and is given by Eq. (11.26). Equating these equations results in the relevant electrical boundary condition

$$D_3(x_3 = 0) = \sum_j \epsilon_{3j}^E E_j + \sum_j \sum_k e_{3jk} S_{jk} \quad . \tag{11.26}$$

Application of these electrical conditions result in a dispersion curve in which the SAW velocity (or coupling strength) is a function of the height of the conductive plane above the substrate. It can be seen that for reasonable interaction with the potential associated with the traveling SAW, the conductive plane has to be placed very close to the surface (see Fig. 11.24).

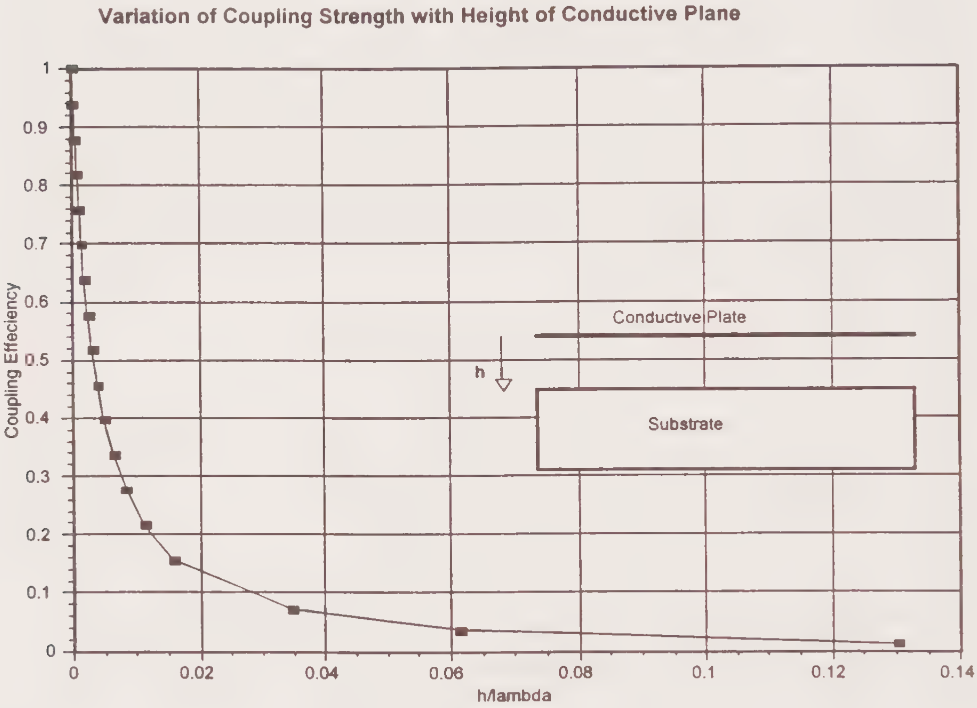


Fig. 11.24. Position of conducting plate.

The reflection coefficient of a single reflector placed above the substrate was then modeled using the concept of approximate field mismatch at the reflector. The results are shown in Fig. 11.25.

As expected this is also a function of the height of the reflector above the substrate. An array of reflectors is needed to ensure that a significant reflection occurs. Figure 11.26 shows the reflection coefficient of an array of reflectors. The number of reflectors used will depend on the sensing capability of the antenna system.

The number of reflectors used will also affect the sensitivity of the device in the following manner. As the reflectors are a part of the seismic mass, any change in the number of reflectors will also change the mass of the seismic mass and hence its displacement due to acceleration. This dependence is shown in Fig. 11.27 for a particular height above the surface.

The theoretical calibration of the sensor is shown in Fig. 11.28. This is for a certain array of configuration. The calibration is linear. Thus by measuring the phase shift the acceleration can easily be detected remotely.

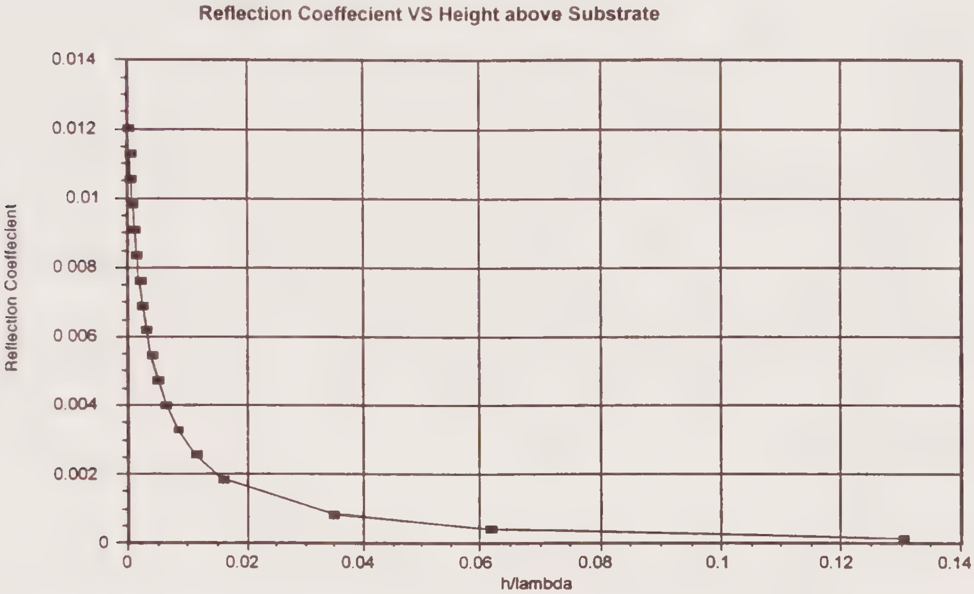


Fig. 11.25. Reflection coefficient vs the height of the conductive structure.

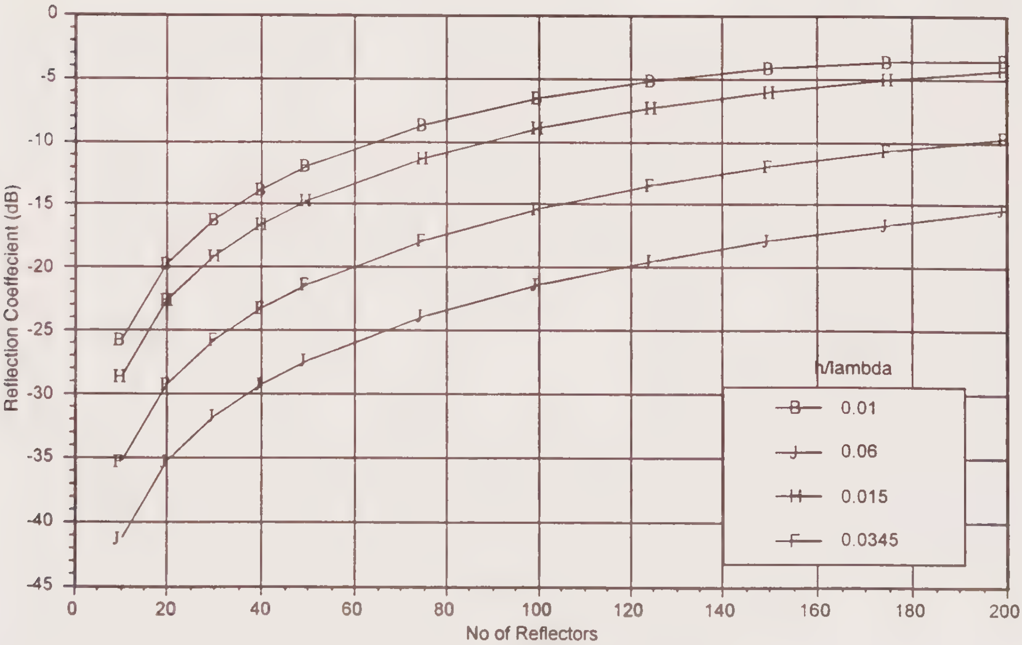


Fig. 11.26. Reflection coefficient vs the number of reflectors.

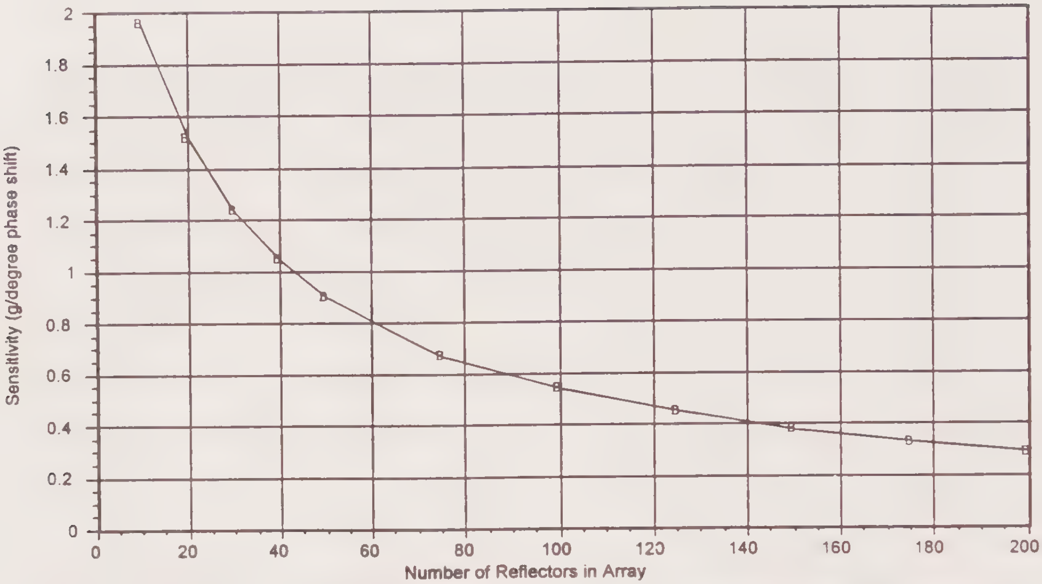


Fig. 11.27. Sensitivity of the accelerometer vs the number of reflectors. Note: The array of reflectors is placed 2000 Å above substrate. Surface velocity is 3475 m/s.

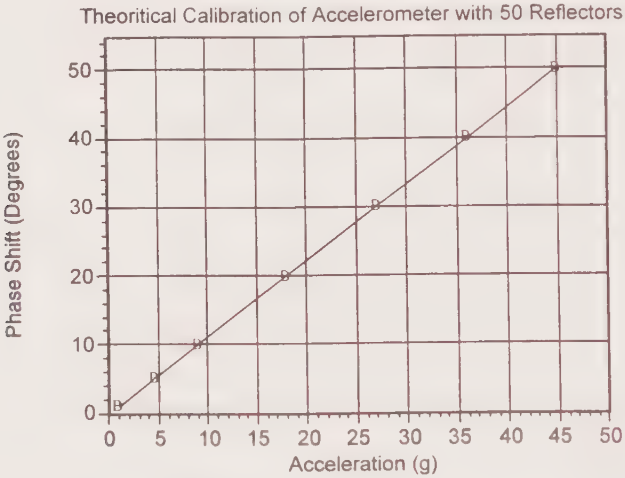


Fig. 11.28. Phase shift vs acceleration.

It can be seen in this design that by varying the number of reflectors the sensitivity can be altered (although a certain minimum number of reflectors are mandatory). The seismic mass can also be altered by choosing a structural material of different density. The choice of these two parameters will allow the design to be modified for different applications.

A less expensive and faster method that is just emerging is the use of photoforming by exposure to UV radiation. The attractive feature of this technique is that electroactive polymer layers can be deposited layerwise in a desired fashion to create various mechanical components at the micron level. As this technology develops, the application of polymer based MEMS to structural health monitoring and process control will be revolutionized.

11.5.3.1 Fabrication method

Lithium niobate is chosen as the SAW substrate on account of its high electromechanical coefficient. The surface of the substrate needs to be highly polished and free of scratches . With this in mind, flip-chip bonding was chosen in order to reduce the handling of the superstrate and hence maintain the performance of the SAW substrate. The IDTs can either be sputtered or etched off the substrate. In either case the fabrication steps involve mask preparation, lithography and etching. The thickness of the metal for the IDTs should be at least 2000 Å in order to make the adequate contact. A very thick layer of metal causes mass loading effects and is detrimental to the device performance. The metallization ratio used for the IDTs is 0.5. The number of IDTs and their aperture is chosen such that the IDTs have an impedance of 50 W.

11.5.4 Suspension Design

The choice of the suspension for the seismic mass determines the linearity of motion and the sensitivity to residual strain. A single support is the simplest and lowest spring constant design, but allows substantial off-line motion and rotation in the suspended mass. Two parallel supports remove the rotation component of the motion, but still introduce off-line motion due to curvature of the beams (arc-length is preserved, vertical distance is not). Two-sided support removes that problem, but greatly increases sensitivity to residual stress.

The seismic mass is fabricated using the principle of sacrificial layer and selective etching. The structural material used is polysilicon. Polysilicon is grown over a layer of sacrificial material (greater than 2000 Å but much less that 3 μm.) The polysilicon is patterned and etched to form the seismic mass. The sacrificial layer (silicon dioxide) is then etched (using HF) to release the free-standing seismic mass. The seismic mass is free to move in the plane parallel to the substrate. The perspective view of the seismic mass is shown in the [Fig. 11.22]

11.5.4.1 Adhesion

After the sacrificial oxide is removed in HF, the wafers are rinsed in DI water and dried. Surface tension of the water under the structures pulls them down to the surface of the wafer, and in some cases causes them to stick (permanently). This problem can be avoided by using thick structural and sacrificial layers, and short structures.

11.5.4.2 Adhesion prevention

There are several other methods of adhesion prevention, which all rely on avoiding the problem of surface tension. These include freeze drying (sublimation) of the final rinsing solution (DI water for Guckel, t-butyl alcohol for Fujita); using integrated polymer support structure during release etching, then ashing in O_2; gradually replacing acetone with photoresist, then spinning and ashing the resist; electrically fusing (vaporizing) micromechanical support structures.

11.5.5 Wafer Bonding

If two wafers with surface oxide layers are hydrated in sulfuric and/or nitric acid and then brought into contact they will stick immediately. If these wafers are subsequently annealed at high temperature, the resulting bond can be essentially perfect. The quality of the bond depends upon the types of oxides used, the temperature, and the cleanliness of the process. If the wafers can be annealed at 1000°C, any common (thermal, PSG, etc.) oxide will do. Since the oxide, rather than the silicon, is involved in the bonding, it is also possible to bond silicon to quartz or other glass substrates. This technique is often used in what is known as a *dissolved wafer process*. This is not possible with due to the sensitivity of the lithium niobate crystal to temperature. There are, however, specialized adhesives available for this purpose. Some of them are listed below with their characteristics.

RTV Adhesives for Mounting SAW Crystals

RTV-102 - General Electric	Room-temperature fast cure, temporary attachment, and corrosive
RTV 3145 - Dow Corning	Room-temperature slow cure, permanent attachment, and noncorrosive
190-3 - Ablestik	Room-temperature moderate cure rate, reduced electromagnetic feedthrough, silver loaded.

11.5.6 Other Design Considerations

In addition to launching surface waves the excited IDTs can also generate bulk acoustic waves. These waves can propagate in any direction within the body of the substrate material. As there in no output IDT in this design, the principle effect of the generation of bulk waves is the reduction of the power available for the generation of surface waves. Hence it is desirable to minimize the generation of bulk waves. The following strategies may be applied to alleviate this situation:

- The bottom surface of the piezoelectric substrate is roughened and coated with a soft conductorlike silver epoxy.
- Use substrate geometries that are not rectangular in shape.
- Choice of the right number of IDT fingers. The input power is converted into bulk-wave energy and surface-wave energy as $P = P_S + P_B$, where P_S

represents the power in the excited SAW wave and P_B the component that is radiated as bulk waves. The ratio of P_S / P_B decreases drastically as the number of finger pairs in the exciting IDT is reduced . For YZ lithium niobate the amount of input power converted into transverse bulk waves increases almost exponentially as the number of IDT fingers is reduced below five.

11.5.7 Acoustic Attenuation

The propagating surface acoustic wave can lose energy in a number of ways, including

1. Viscous damping in crystal lattice of the piezoelectric substrate , through which the SAW energy is dissipated as a minuscule amount of heat.
2. Acousto-electric losses caused by resistive films on lithium Niobate or special doping of semiconductors
3. Crystal imperfections.

In vacuum SAW energy loss is approximately proportional to the square of the operating frequency at or near room temperature. There will be a small ultrasonic loss contribution when the SAW surface is surrounded by air or some inert atmosphere due to coupling of some ultrasonic energy to the surrounding atmosphere.

The insertion loss contribution due to acoustic attenuation will be in addition to those due to mismatching, diffraction beam steering and IDT ohmic loss. For YZ lithium niobate the acoustic loss is given by $0.88F^{1.9} + 0.19\,F$ where F is the frequency in GHz.

11.5.8 Advantages of the Design Presented

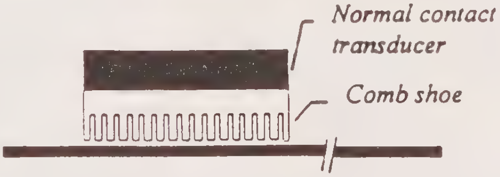
1. Electromagnetic feedthrough (crosstalk) occurs when the input and the output IDT effectively form a capacitor so that the electromagnetic energy can be directly coupled from the input IDT to the output IDT. The susceptance of the equivalent series capacitance between the IDT increases with frequency, so will the severity of the feedthrough. This problem is avoided in the current design as there is only one pair of IDTs.
2. Since the handling of the SAW substrate is kept to a minimum , a better quality of the SAW is expected.
3. The MEMS has no power supply of its own.
4. It can be remotely addressed. Since it operates at 1 GHz it is compatible with existing antenna systems.
5. The device parameters such as the mass and aperture of the IDT can be changed to alter the sensitivity to acceleration and for impedance matching.

Calculations governing the sensing capability of the accelerometer:

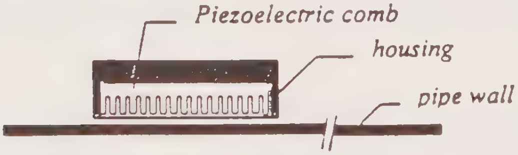
1. Phase shift (Q) is given by
   $Q = 2\, P f t$ radians where f is the frequency
   $Q = 360\, f t$ degrees
   For $f = 1$ GHz , $Q = 360 \times 10^{-9}\, t$
   1° phase shift is equivalent to a time delay of 2.778 ps.
2. The minimum distance to be moved to result in 1° phase shift @ 3000 m/s (speed of the SAW) = $3000 \times 2.778 \times 10^{-12} \times 0.5$
   $= 2.0835 \times 10^{-9}$ m or 2.0835 nm
3. Considering the beam to be clamped at both ends, the maximum deflection at mid span is given by $\delta = \dfrac{WL^3}{192EI}$, where $W = m.a/2$.
   For $m = 0.16$ µg, $E = 1.7 \times 10^{11}$ N/m², $L = 575$ µm , $b = 2$ µm, $d = 2$ µm, deflection $= 3.4946 \times 10^{-10}\, a$.
   For an acceleration of 1g , deflection = 3.428 nm or 1.645° phase shift.
   For a deflection of 2.0835 nm or 1° phase shift an acceleration of 0.60775 g is required.
   The sensitivity of the accelerometer is 0.60775 g per degree phase shift.
   The maximum deflection allowable before failure is 1.6 µm*. This dictates the maximum measurable acceleration, which comes out to be 46.67 g.
4. Acoustic wavelength for a SAW at 3000 m/s and 1GhZ frequency is 3 µm. Spacing between the fingers of the IDT is therefore 1.5 µm. Width of each IDT is also 1.5 µm. This is the condition for maximum efficiency.
5. Length of reflectors are 100–200 wavelengths and the distance between the IDT and array of reflectors are 0.4mm

11.6 SENSORS FOR HEALTH MONITORING

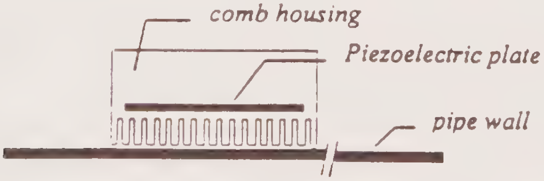
Aircraft and spacecraft structures require an extensive amount of health monitoring because their safety is critical. Health monitoring is performed with the sensors conveniently bonded on the structure's surface. Even though the strain, vibration, etc., can be measured using the surface sensors, it is important to study the internal damages, cracks, etc. IDT and MEMS-based sensors are being developed for such purposes. The IDT microsensors described before can generate surface waves and bulk waves. The bulk waves travel into the structure and are scattered by the presence of cracks, voids, etc. The scattered signal will then be analyzed by the sensors.[42] Another type of sensor known as a comb transducer, which can be made with the photolithographic process, is currently being used for nondestructive evaluation and health monitoring of aging aircraft.[43] The comb transducer can be made in different ways. As shown in Fig. 11.29 (a), the comb transducer can be a comb shoe attached to the front face of a normal beam transducer. The normal beam transducer drives all of the elements to vibrate in phase so as to excite guided waves. As in Fig. 11.29(b), a piezoelectric plate is directly placed in a housing with the comb at the bottom. Similar to Fig. 11.29(a), the piezoelectric plate drives the comb to vibrate in phase. The comb transducer in Fig. 11.29(c) consists of a group of independent elements that are equally spaced arranged in parallel order. An ultrasonic pulsar/receiver sends electrical pulses in phase to all elements to excite guided waves as transducers in (a) and (b). The generated guided wave on a plate may be received using three methods as illustrated in Fig. 11.30. Figure 11.30(a) shows the pulse echo method: the comb transducer serves as both sender and receiver, and the echo is received by the comb. Figure 11.30(b) is the through transmission method using a comb transducer to receive the signal. Both Figs. 11.30(a) and 11.30(b) receive a particular selected wave mode because the comb transducers are designed for it. Figure 11.30(c) is the through transmission method using an angle beam transducer to receive the signal. This method is also very effective in receiving the particular excited mode because the Snell laws allow a particular mode conversion to a longitudinal wave that is received by the angle beam transducer.


Normal contact transducer
Comb shoe

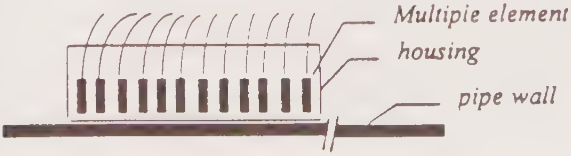
(a) Transducer on top of comb shoe


Piezoelectric comb
housing
pipe wall

(b) A piezoelectric comb in a comb housing

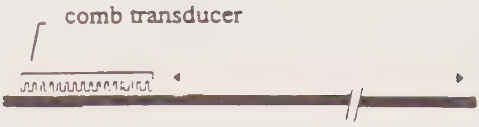

comb housing
Piezoelectric plate
pipe wall

(c) A piezoelectric plate in a comb housing

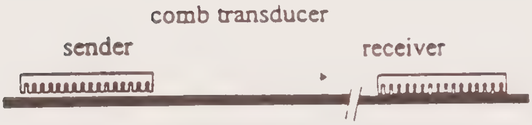

Multiple element
housing
pipe wall

(d) A multiple element comb transducer

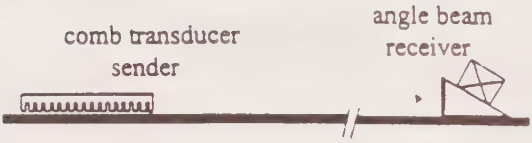
Fig. 11.29. Configuration possibilities of comb transducers.



(a) pulse echo



(b) comb/comb through transmission



(c) comb/angle beam through transmission

Fig. 11.30. Experimental setups using the comb transducer technique.

11.7 APPLICATIONS TO SMART STRUCTURES

11.7.1 Deflection and Strain of Flex-Type Beam Structures

In this section a remote SAW strain sensor system is employed to study the deflection and strain of a "flex-beam" type structure of a helicopter. The system consists of remotely readable passive SAW sensors and a microwave reading system. The experimental results confirm the theoretical analysis using a simple model and show the feasibility of using passive SAW sensors to measure structural strains.

The microwave reading system used in this system is actually a FM radar. The FM signal send by the system antenna is expressed as

$$S(t) = A\cos(\omega_0 + \mu t/2)t \quad , \tag{11.27}$$

where ω_0 is the start frequency of the FM signal, μ is the rate of the modulation, and t is time.

The echoes from the reflectors $S_1(t)$ and $S_2(t)$ are the same as the transmitted signal $S(t)$, but with times delays t_1 and t_2 respectively. They are written as

$$S_1(t) = S_1 \cos(\omega_0 + \mu t/2)(t - t_1) \tag{11.28}$$

$$S_2(t) = S_2 \cos(\omega_0 + \mu t/2)(t - t_2) \quad , \tag{11.29}$$

where

$$t_1 = 2d_1/v + \tau_e \tag{11.30}$$

$$t_2 = 2d_2/v + \tau_e \quad , \tag{11.31}$$

where v is the SAW velocity, d_1 and d_2 are the distances from the IDT transducer to the reflectors respectively, and τ_e is the total of other delays such as electromagnetic wave traveling time or delay in electronic circuit and devices, which is the same for both echoes.

Through the mixer, which uses the transmitted signal as reference, and low-pass filter, frequency differential signals are obtained as

$$\begin{aligned} E_1(t) &= E_1 \cos[\mu t_1 t + (\omega_0 t_1 - \mu t_1^2)] \\ &= E_1 \cos[\omega_1 t + \phi_1] \end{aligned} \tag{11.32}$$

$$\begin{aligned} E_2(t) &= E_2 \cos[\mu t_2 t + (\omega_0 t_2 - \mu t_2^2)] \\ &- E_2 \cos[\omega_2 t + \phi_2] \end{aligned} \quad . \tag{11.33}$$

respectively. The frequencies and phases of these two signals both depend on the delay times. The two signals can be separated in the frequency domain. Since the ω_0 is usually greater than m, the phase shift is more sensitive to the variation of delay time than the frequency.

The difference of the two phases can be written as

$$\phi = \phi_1 - \phi_2 = [\omega_0 - \mu(t_1 + t_2)/2](t_2 - t_1) \quad , \tag{11.34}$$

where the extra delay time of the second echo referring to the first is equal to the round trip time of the acoustic wave traveling from the first reflector to the second and is presented as

$$\tau = t_2 - t_1 = \frac{2d}{v} \quad ,$$

where d is the distance between the two reflectors.

The phase difference is sensitive to the change of the delay times. The variation of the phase difference due to the change of the delay times is expressed as

$$\Delta\phi = [\omega_0 - \mu(t_1 + t_2)/2]\Delta\tau \quad . \tag{11.35}$$

Since ω_0 is usually much larger than the item of $\mu(t_1 + t_2)/2$, then we have

$$\Delta\phi = \omega_0 \Delta\tau \quad .$$

The wave traveling time t is proportional to the distance between the two reflectors and inversely proportional to the velocity. If neglecting the possible velocity variation of the SAW under strain and taking into account only the direct effect of the distance change, we have

$$\Delta\phi = \omega_0 \frac{2\epsilon d}{v} = \omega_0 \epsilon \tau_0 \quad , \tag{11.36}$$

where is e is strain and τ_0 is the traveling time when the strain is zero. The sensitivity of this remote sensor system depends on operating frequency and the round traveling time of the SAW between the two reflectors.

11.7.1.1 Experimental setup and results

The test setup is shown in Fig. 11.31. The SAW wafer package is glued to a fiberglass cantilever beam 172 mm long, 25 mm wide, and 1.5 mm thick. The center of the SAW wafer is 53.3 mm from the root of the beam and the SAW propagating direction is along the beam length. If we neglect the effect of the boned SAW wafer, the strain on the surface of the beam can be calculated by simple beam theory

$$\epsilon_x = -d_s \frac{3t_b(l-x)}{2l^3} \quad , \tag{11.37}$$

where l is the length of the beam, x is the distance of the location from the root, t_b is the thickness, and d_s is the displacement at the tip of the beam. In the experiment, the beam is bent by a force at the tip, and the tip displacement is measured by a dial indicator.

The remote SAW sensor system successfully detected the shift of the phase difference of the two echoes with the variation of the tip displacement. The results are presented in Fig.11.32 in the form of the shifts vs strains. The strains are calculated according to the tip displacements by Eq. (11.37).

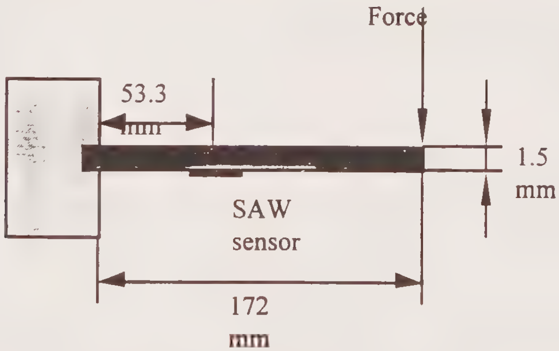


Fig. 11.31. Passive SAW sensor on test cantilever beam.

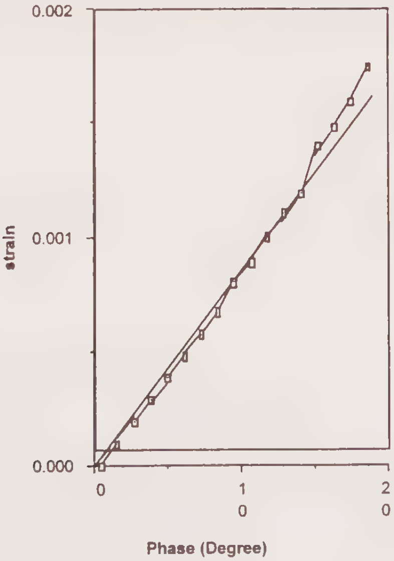


Fig. 11.32. Experimental results of strain versus phase change.

The data show that the phase shift varies with the strain linearly when the stain is less than .0012. For larger strains, nonlinearity and larger fluctuation appear. The nonlinearity may be due to the glue used to bond the SAW package, the large deformation of the beam, or other causes. The maximum tip displacement in the test is 34 mm. It is compatible with the length of the beam and beyond the range where simple beam theory is valid.

The sensitivity and resolution converted from the measurement data is about $1.09 \cdot 10^4$ degrees and $0.92 \cdot 10^{-4}$, respectively. They are lower than the theoretical estimation of $3.3 \cdot 10^4$ degrees and $3 \cdot 10^{-5}$. Since the thickness of the SAW wafer and the plastic case is not negligible, compared with the thickness of the cantilever beam in the test, it is expected that the strain on the wafer surface is less than the strain on the surface of the cantilever beam when the SAW sensor is absent. The latter is used to calculate the sensitivity and resolution.

Considering this factor and that the softness of the plastic case also may decrease the sensitivity of the inside SAW sensor, the test results reasonably confirm the theoretical estimation.

11.7.2 Dynamic Strain

In actuality, rotorcraft blades undergo vibration and deflection. Many times the blades also carry a weight at the tip to control certain harmonics. Here we simulate such structures as a simple cantilever beam subjected to dynamic loads, and the strain is measured by IDT sensors through the phase angles. The measured phase angles for the unweighted and weighted cantilever beam undergoing vibration at the first resonance are presented in Fig. 11.33. The weighted beam has been shifted down one degree so that the two graphs can be compared. This extra degree was a result of the added static strain in the beam due to the addition of the weight. The phase change for both beams ranges over a phase angle change of four degrees. The weighted beam may be experiencing a greater range of phase angles, but the sensor does not have enough sensitivity to measure this difference. The effect can be noted at the peaks of the oscillation where two adjacent data points have equal phase angles. It is reasonable to assume that during the time in between these points, the phase angle would have further increased.

An interesting phenomenon can be noted by observing the period between successive peaks of the phase angle. The unweighted beam actually had a longer period of oscillation than the weighted beam, as noted in Fig. 11.33. This is counterintuitive, but may be attributed to the different excitation frequencies for each beam. The excitation frequencies differ because each beam was vibrated at its own first mode of resonance.

This information can be combined with the relationship between phase angle change and strain from Fig. 11.32 to yield the train in the beams over time. For this plot (Fig. 11.34), an angle of 13 degrees was taken to be the midpoint of the motion and at zero strain due to the dynamic loading.

The concept presented above has been implemented on rotating blades. First the strain and the deflection were measured remotely by the antenna communicating with the sensors (via the built-in antenna) while the blades were stationary and then when the blades were rotating. The rotational speed is close to that of an actual helicopter. The results presented in Fig. 11.35 are encouraging for further research.

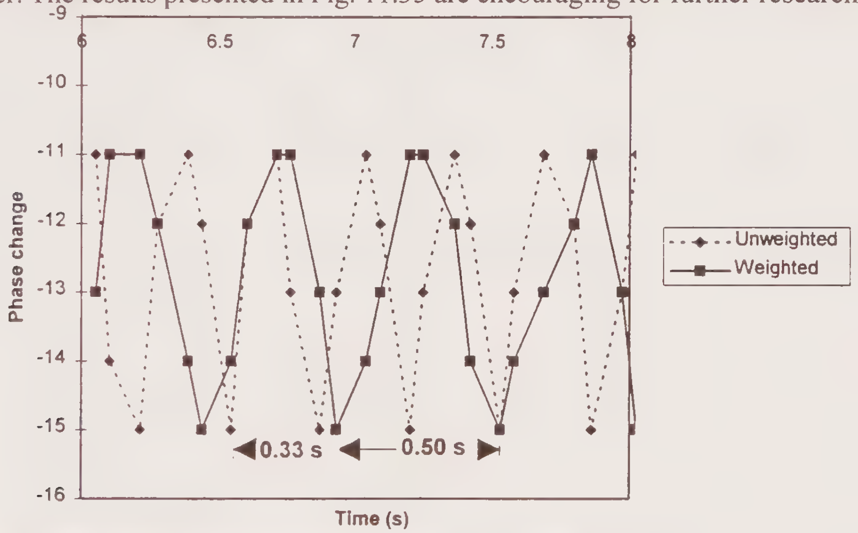


Fig. 11.33. Phase change vs. time for weighted and unweighted beams .

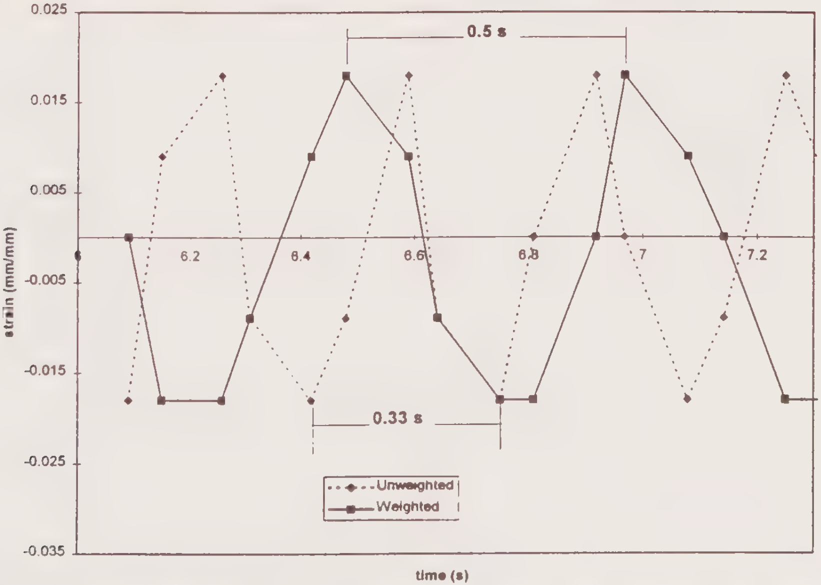


Fig. 11.34. Strain measured by SAW sensor over time for a weighted and unweighted vibrating beam.

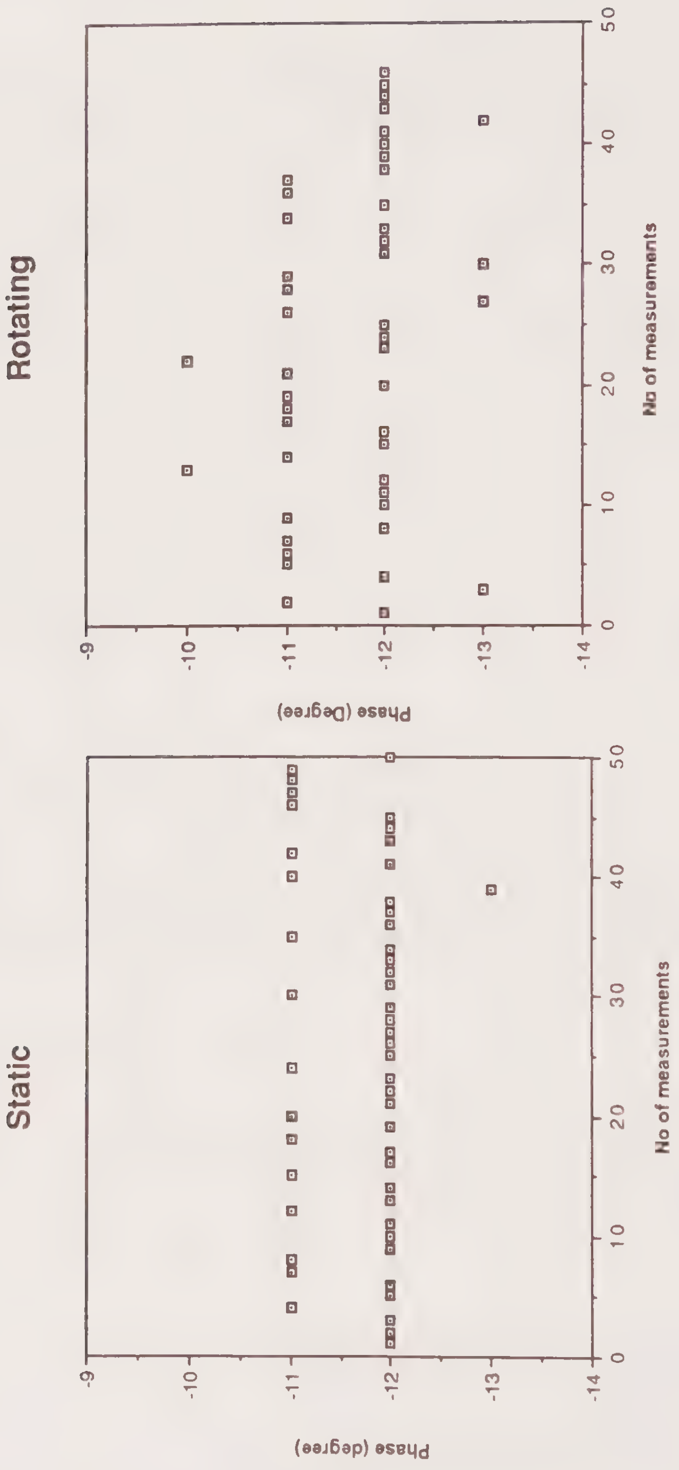


Fig. 11.35. Phase shift and hence the strain measurement of stationary and rotating blades.

11.7.3 Local and Global Sensing and Control of Vibration and Drag in Aircraft

In this case the sensors described above are used to measure both pressure and shear of the fluid flow on aerospace structures. The waveform measurements may be monitored at a remote location either at the cockpit or elsewhere via the antennas in the sensors and an outside antenna. The integrated MEMS actuators with suitable smart electronics etched onto the structure are controlled by the built-in antennas through feedback and feedforward control architecture.

The integration of such materials and smart electronics into the skin of airfoil is ideal for sensing and controlling vibration and drag. The basic idea of this concept involves detection of the point of transition from laminar to turbulent flow and transmitting acoustical energy into the boundary layer so that the low-energy fluid particles accelerate in the transverse direction and mix with the high-energy flow outside of the boundary layer.

A schematic arrangement that connects SAW and MEMS devices to antenna onto the airfoil of aircraft is shown in Figs. 11.36 and 11.37.

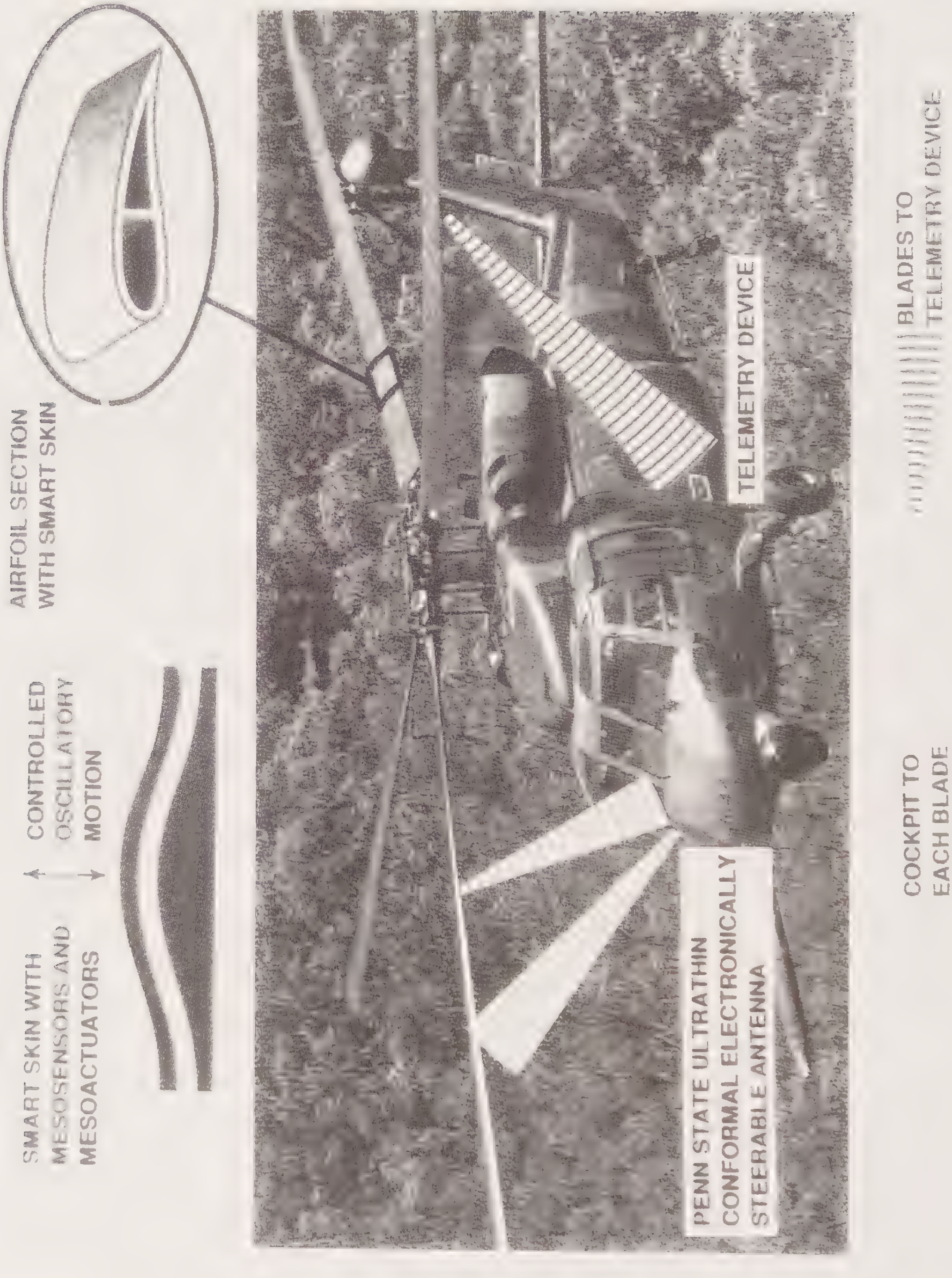

SMART SKIN WITH MESOSENSORS AND MESOACTUATORS
CONTROLLED OSCILLATORY MOTION
AIRFOIL SECTION WITH SMART SKIN
PENN STATE ULTRATHIN CONFORMAL ELECTRONICALLY STEERABLE ANTENNA
TELEMETRY DEVICE
COCKPIT TO EACH BLADE
BLADES TO TELEMETRY DEVICE

Fig. 11.36. Helicopter with built-in sensors and electronically steerable antenna.

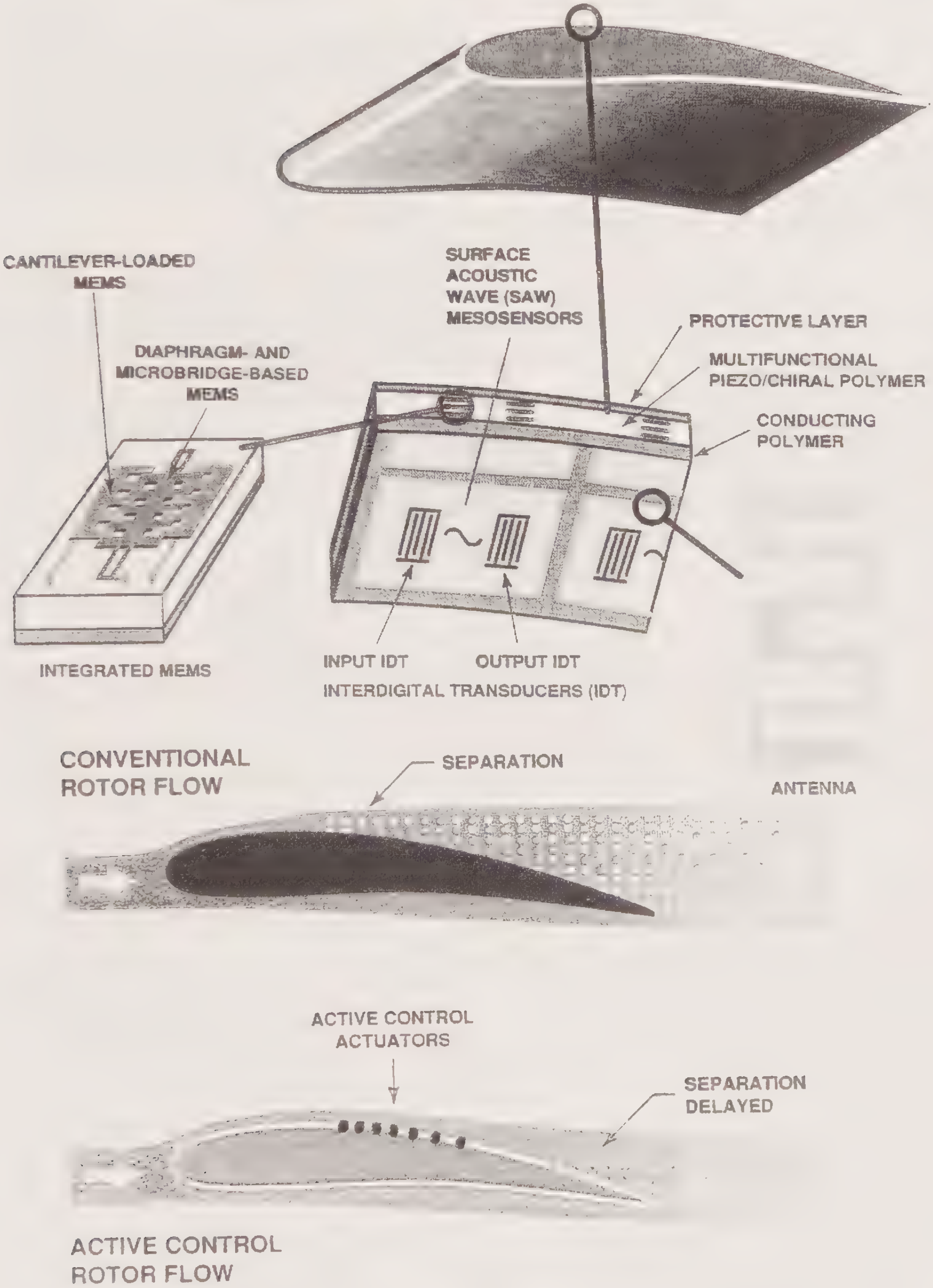
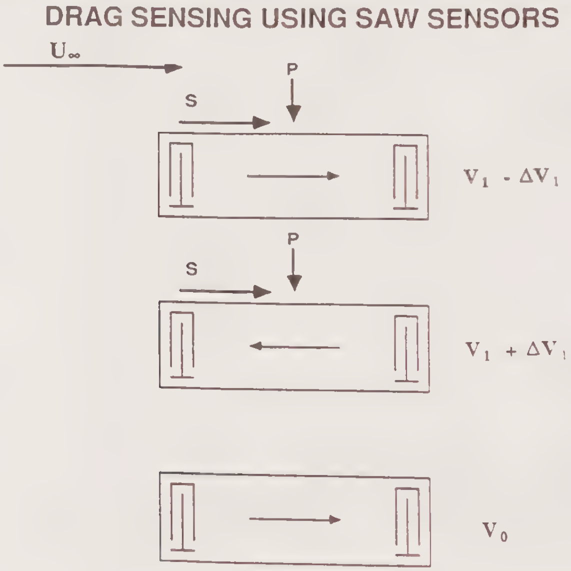

CANTILEVER-LOADED MEMS
SURFACE ACOUSTIC WAVE (SAW) MESOSENSORS
PROTECTIVE LAYER
MULTIFUNCTIONAL PIEZO/CHIRAL POLYMER
DIAPHRAGM- AND MICROBRIDGE-BASED MEMS
CONDUCTING POLYMER
INTEGRATED MEMS
INPUT IDT
OUTPUT IDT
INTERDIGITAL TRANSDUCERS (IDT)
CONVENTIONAL ROTOR FLOW
SEPARATION
ANTENNA
ACTIVE CONTROL ACTUATORS
SEPARATION DELAYED
ACTIVE CONTROL ROTOR FLOW

Fig. 11.37. Airfoil section with MEMS and actuators.

The smart skins derived from this integration are ideally suited for drag sensing and control of aircraft and spacecraft. One can also induce microriblets on the smart skin for drag reduction using micromachining techniques. These microriblets can be actively produced by the MEMS techniques at the desired location. The detection and control of noise and vibration as caused by both external and internal forces may be countered through the use of polymeric piezosensors and actuators.

For sensing drag pressure and shear, SAW MEMS devices are used as described below. In principle, two SAW sensors are normally employed as shown in Fig. 11.38; in one of the SAW devices the waves propagate in the same direction of the air flow while in the other the waves propagate in the opposite direction. Using the velocity V_0 as the reference when there is flow, we could relate the pressure and the shear from these two SAW sensors as shown in Fig. 11.38.

The sensors arranged in the rosette form can provide the direction of the flow (see Fig. 11.39). A cantilever-beam SAW force sensor attached to a floating thin and flat membrane will enhance the stress many orders of magnitude (see Fig. 11.40).



$$\frac{(V_1 + \Delta V_1) + (V_1 - \Delta V_1)}{2} - V_0 = V_1 - V_0 = \Delta V \propto P$$

$$(V_1 + \Delta V_1) - (V_1 - \Delta V_1) = 2\Delta V_1 \propto S$$

Fig. 11.38. SAW pressure and shear sensor.

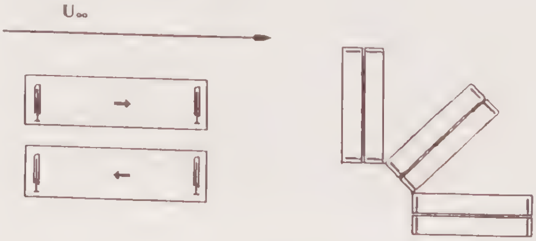


Fig. 11.39. Rosette arrangement of SAW sensors.

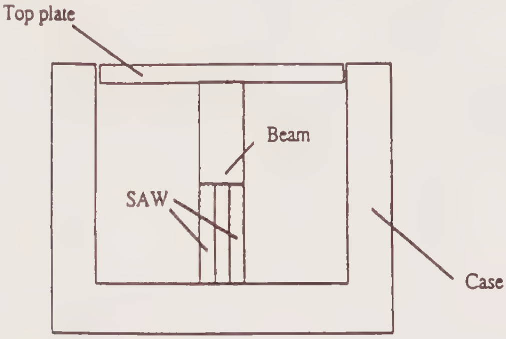


Fig. 11.40. Cantilever beam SAW sensor.

The device can be flush mounted in an aircraft. The composite cantilever beam has two SAW devices fabricated by using lithography. The active areas of the two devices, where surface acoustic waves propagate, face opposite sides. The moment applied by the attached membrane exposed to shear, one active surface is in tension and the other is in compression. Dual sensor approach also compensates for any changes in other physical variables such as temperature or pressure. A preliminary experiment was performed at Penn State Wind Tunnel. The tunnel is a 16-inch low-speed unsteady boundary layer wind tunnel. The sensor system was mounted on a horizontal internal plate that divides the tunnel of rectangular cross section into two equal parts. The sensing membrane of the sensor was in the up position. The velocity of the wind in the tunnel was monitored by a pressure sensor close to the air inlet. The output of the pressure sensor is voltage, which is proportional to $P = m\,(V - V_0)$ with $m = 66.89$ Pa/V, where V_0 is the velocity of the surface wave.

A calibrated curve of the wall shear stress t and pressure P is also available for the tunnel. The shear stress is proportional to the pressure P, $t = kP$ where k is approximately equal to 0.003129 for the wind tunnel under consideration. First, the sensor is calibrated and the relation between the pressure and shear vs the phase shift is determined. In the experiment, the wind speed was increased step by step to a maximum first and then decreased gradually. The output frequency of the SAW sensor and time was recorded. The results are shown in Figs. 11.41 and 11.42. The sensor system will be implemented in model airfoil and the performance will be assessed later.

The concept of microriblets for drag reduction has been presented by Huang and Ho[44] and by the authors.[45, 46] Using a 3D MEMS structure as presented by Varadan,[11] one could produce active microriblets to control the drag. Even a 5% drag reduction will result in an enormous fuel savings.[47]

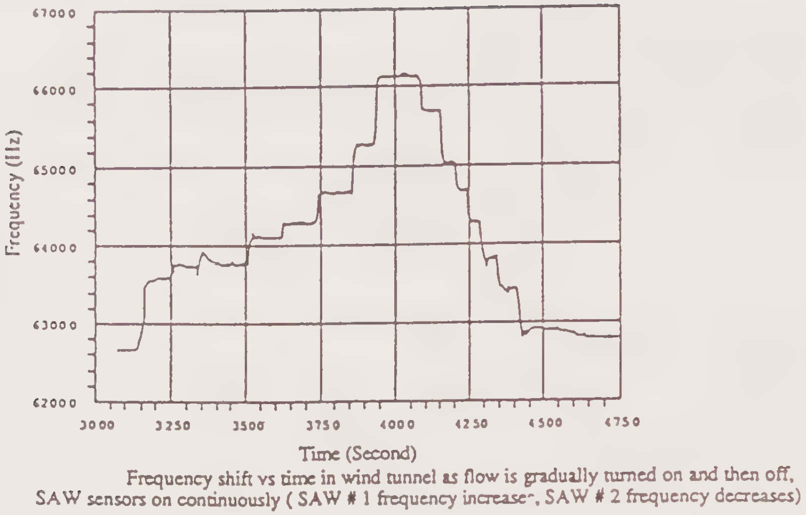


Fig. 11.41. SAW results in wind tunnel.

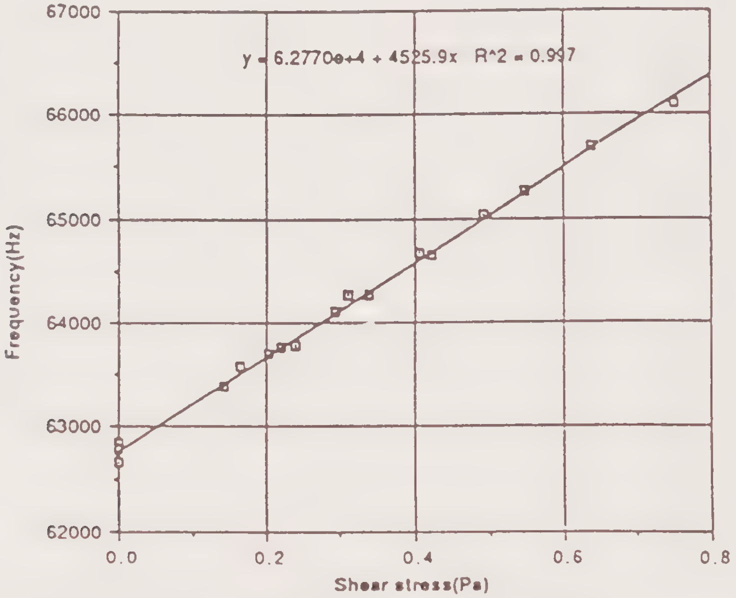


Fig. 11.42. Frequency shift vs. wall shear stress using calibration.

11.7.4 "Smart Plate" for Noise Reduction using MEMS Technology

Active structural acoustic control has attracted growing attention in recent years. Piezoelectric sensors and actuators are widely used because they can be embedded in structures or bonded to surfaces with minimum interference to the structure design. Much research has been devoted to sound transmission through plates or shell.[48-50] A theoretical analysis by Fuller showed the feasibility of actively controlling the sound transmission/radiation with a few actuators applying forces to the plate.[48] The concept was proved by later experiments.[49] The analytical work aimed at applying the technique to a cylindrical shell such as an aircraft fuselage also shows encouraging results.[50] The possibility of using on-plate error sensors to achieve significant global reduction has been addressed.[51] Most of these investigations deal with the cases of transmission between two free acoustic fields or into cylindrical enclosures, and the boundary conditions involved are ideal, such as simple support.

The present experimental investigation addresses the active control of sound transmitted from an enclosure through plates by using on-plate MEMS sensors and actuators.[52, 53] This experiment will provide the foundation for cabin noise reduction in aircraft. In cabin noise reduction, the problem becomes simpler since we want to control the noise coming from outside to the interior of the cabin. The plate is mounted at an opening of a cubic sound enclosure. The boundary condition of the plate is not ideal; it is rather close to the mounting in most practical applications. The sound transmitted through the plate is measured by a microphone array that scans a hemispherical surface. Measurements show that the spectrum of the transmitted sound has peaks that coincide with the resonance of the plate. The target of the active control is first to reduce these transmission peaks that result in poor isolation performance. This strategy requires a relatively simple, low-cost control system and may be applied to a wide variety of structures such as machine cases, windows, etc. The capability of a one-sensor one-actuator control system is evaluated. Significant reductions of 15–22 dB are obtained by use of this simplest control system at the resonance frequencies of the first three symmetric modes.

In this study, an aluminum square plate is used as a test sample. The dimensions of the plate are 14" × 14" × 0.015". The MEMS sensor fabricated is 400mm × 400 mm. Its performance is compared with a regular PZT sensor and actuator as shown in Fig. 11.43. Three PZT (Lead zirconate titanate) pieces are bonded on the plate by epoxy as shown in Fig. 43. One is used as an actuator, indicated as A1 in the figure. Those marked as S1 and S2 are sensors. The geometry of A1 and S2 are the same. They are disks with a diameter of 1.5" and a thickness of .02". S1 is 1" in diameter. Actuator A1 is located at the center of the plate, sensor S1 at the middle from the center to an edge, and sensor S2 at the middle from the center to a corner. To compare the performance of the MEMS sensor, it is located on the actuator at the center of the plate.

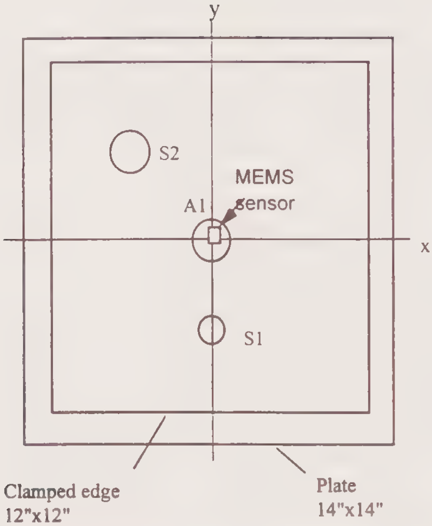


Fig. 11.43. The aluminum plate with piezoelectric actuator and sensors and MEMS sensor.

Since the actuator is located at the center of the plate, we can expect the actuator to have good couplings to most symmetric vibration modes such as (1, 1), (3, 1), and (1, 3). The symmetric modes are usually more efficient in sound radiation as compared with the antisymmetric ones, especially in low-frequency ranges.

The plate is mounted on a five-sided wood box in a semi-anechoic chamber as shown in Fig. 11.44. The box measures 12" × 12" × 12" inside. The walls of the box are made from thick (1") oak plates, and fiber-rubber-fiber acoustic isolation mats are put in front of the walls to prevent the sound in the box from being transmitted through the walls. The plate is clamped on the edges of the opening of the box by a steel frame. The movable area of the plate is 12" × 12". A soft rubber gasket is placed between the plate and the wood. The boundary condition is somewhere between the clamped and the simple support. A loudspeaker generates the sound field in the enclosure. The sound level outside is measured by a microphone array. The eleven elements of the array are uniformly distributed in a half circle in a distance of 15" from the center of the plate. The array is able to rotate in 180° to cover a hemispherical surface above the plate and the baffle. The baffle is a 0.75"-thick plywood board.

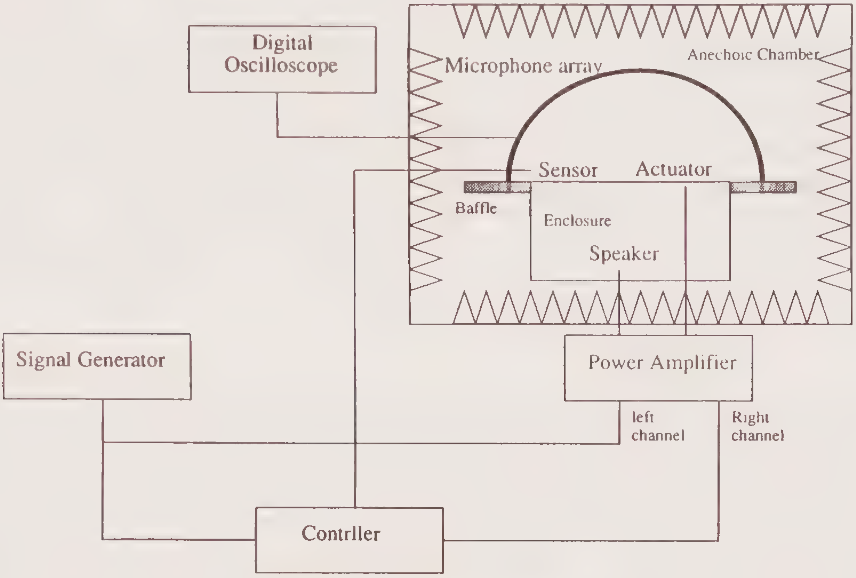

Digital Oscilloscope
Microphone array
Anechoic Chamber
Sensor
Actuator
Baffle
Enclosure
Speaker
Signal Generator
Power Amplifier
left channel
Right channel
Contrller

Fig. 11.44. Noise control and measurement.

The global power reduction obtained by active control may be defined as

$$R = 10\log(W_o/W_c) \quad , \tag{11.38}$$

where W_o and W_c are the sound power radiated from the original and controlled plate respectively. Although the measuring surface in this experiment is not far enough to satisfy the far-field condition, we estimate the global reduction for convenience. Approximately, we have

$$R = 10\log\left[\int_{S_h}(|p_o|^2)ds\Big/\int_{S_h}(|p_c|^2)ds\right] \quad , \tag{11.39}$$

where S_h is the hemispherical surface and p_o and p_c are the original and control pressures, respectively.

Actually, the sound pressure is measured at discrete points. The integrals are replaced by the sum of the multiplication of the square of the sound pressure and the surrounding area.

First, the original plate without active control was investigated. In the measurement, a white noise with a bandwidth from 20 Hz to 2 kHz was applied to the loudspeaker. The sound pressure at the position 15" up to the center of the plate and the outputs of sensor S1 and S2 and that of the MEMS sensor were recorded.

The performance of the one-sensor/one-actuator system was tested in this experiment. The signal from sensor S1 or S2 was taken as an error signal by the controller. The controller adjusted the amplitude and phase of the voltage applied to the actuator in such a way as to make the error signal minimal, while the voltage applied to the speaker remained constant. In this experiment, it was usually more than 40 dB less than the original. We chose S1 as the basic sensor because the output appears to be more sensitive to the modes with stronger sound radiation and less sensitive to the inefficient radiating modes. In the case where sensor S1 did not result in a reasonable sound control, sensor S2 was tested.

The best result was obtained at the first resonance frequency of 79 Hz. Sensor S1 was used as the error sensor. The reduction by the active control is shown in Fig. 11.45. The original sound field on the hemispherical surface was quite uniform. The variation was about ± 1.5 dB. The active control reduced the sound level at all the measuring positions on the hemispherical surface by 20 to 26 dB. The global reduction was 22 dB. The same kind of result was also obtained using a MEMS sensor at this frequency.

Comparisons were then made for other frequencies using PZT and MEMS sensors as shown in Figs.11.45 and 11.46. It is to be noted that the MEMS sensor performance is comparable and sometimes better than the conventional large-size PZT sensor.

The use of MEMS technology for cabin noise reduction using the multifunctional polymer mentioned as "smart wall paper" will be reported later.

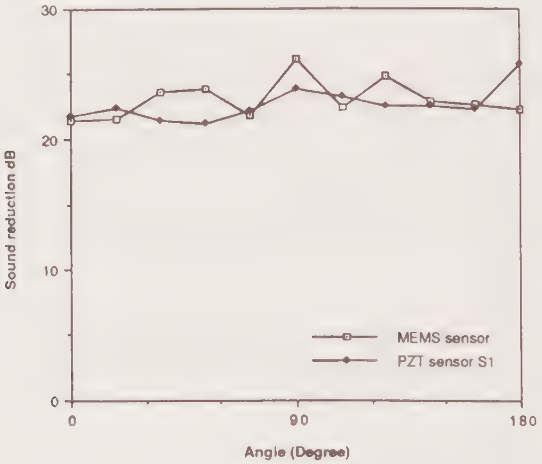


Fig. 11.45. Comparison of MEMS and PZT sensors at 79 Hz.

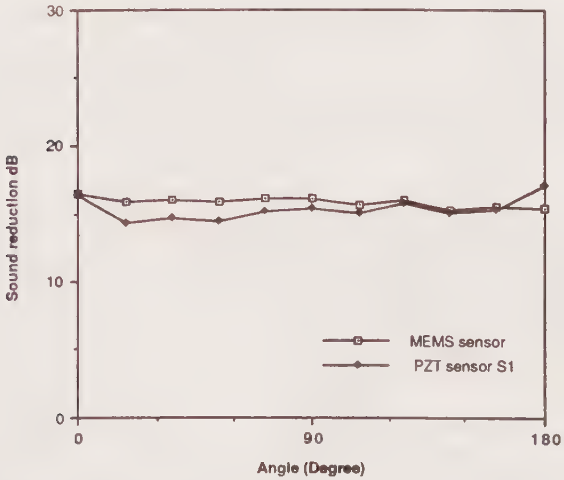


Fig. 11.46. Comparison of MEMS and PZT sensors at 257 Hz.

11.7.5 Health Monitoring of Aircraft Structures

The theoretical and numerical challenges in simulating a structure with embedded sensors using comb transducers, MEMS, SAW devices, etc., are many. The material may be inhomogeneous; the piezoelectric phases are anisotropic and involve coupled elastodynamic and electric fields; the structure may have protective coatings; the structure may be exposed to a semi-infinite environment (liquid). A numerical simulation method that can be applied to arbitrary geometry, material constituents, coupled field problems, and transient effects calls for techniques such as hybrid finite element methods.

The authors have developed several versatile codes that can look at optimization of sensor design, sensor placement, and antenna interfaces. Useful sensor parameters such as electrical conductance and susceptibility, crosstalk between sensor elements, sensor fidelity and efficiency, sensor/structure interface problems, and effect of material damping can all be calculated and studied numerically. Simpler calculations for the design and performance simulation of various designs for MEMS and SAW devices as well as transducer arrays and comb transducers can also be done in a much shorter time at much less cost. There is a well-established precedent for the finite element and boundary element modeling (FEM and BEM) of structures with embedded sensors.[55] This is now interfaced with new models that have been developed for various types of transducers. Transient problems can also be handled using the unconditionally stable α method for time stepping the dynamical equations of the system.

It is generally well known in the NDE community that inspection redundancy can often improve the overall probability of defect detection. In addition to bulk wave, surface-wave and guided-wave ultrasonic NDE techniques, we are currently pursuing research on a multitechnology program including MEMS devices and microwave NDE. The overall NDE methodology and decision algorithm development program is illustrated in Fig. 11.47. The application to aircraft structures is given in Figs. 11.48 and 11.49. In Fig. 11.49 a combination of fiber optics and antenna technology is given for future health monitoring of structures.

In Fig. 11.50, an optical carrier wave with microwave signal is fed through an optical fiber. Using a pin-photodiode, the microwave signal is transmitted into the structure containing cracks via an antenna. The signal is scattered by the crack and the scattered signal is received by the antenna. A three dimensional image of the flaws can then be constructed using the ramp response signature as shown in Ref. 54.

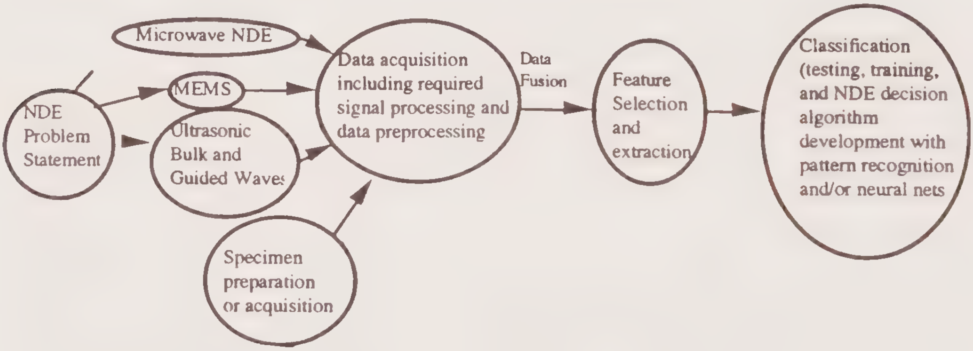

Microwave NDE
MEMS
NDE Problem Statement
Ultrasonic Bulk and Guided Waves
Data acquisition including required signal processing and data preprocessing
Data Fusion
Feature Selection and extraction
Classification (testing, training, and NDE decision algorithm development with pattern recognition and/or neural nets
Specimen preparation or acquisition

Fig. 11.47. Decision algorithm program.

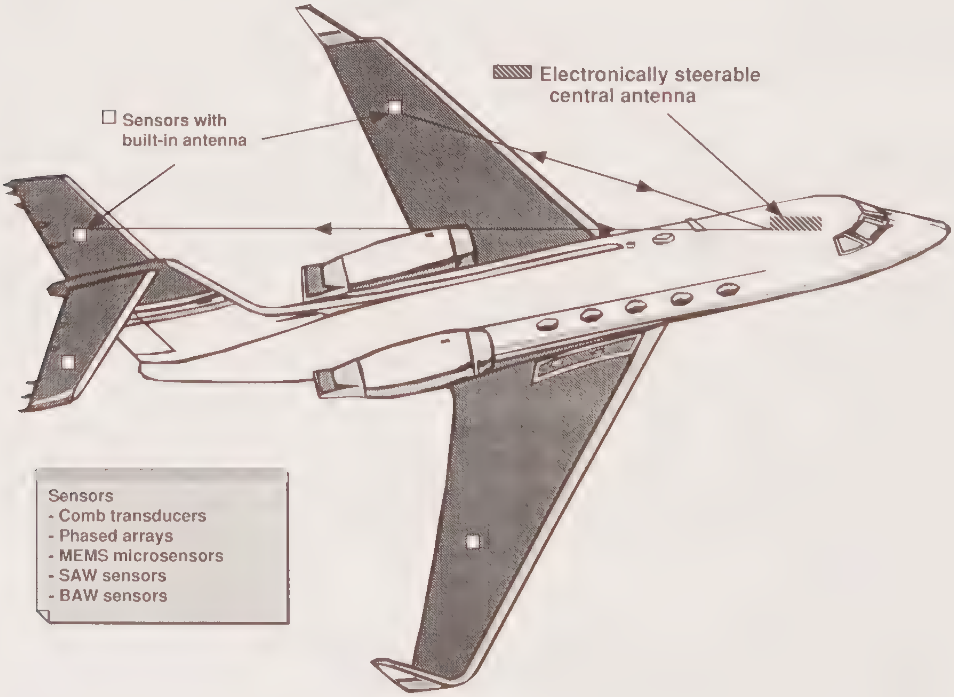

▨ Electronically steerable central antenna
☐ Sensors with built-in antenna
Sensors
- Comb transducers
- Phased arrays
- MEMS microsensors
- SAW sensors
- BAW sensors

Fig. 11.48. Wireless telemetry for health monitoring of aircraft.

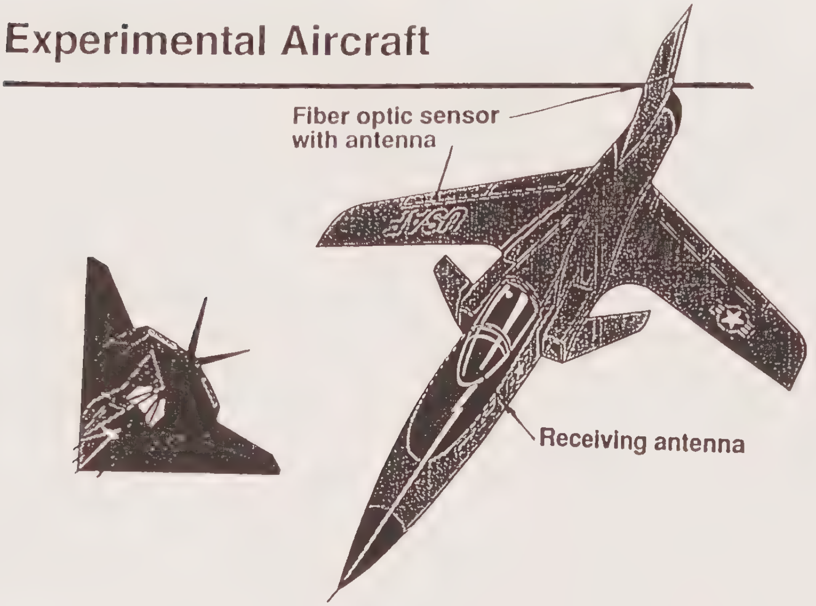

Experimental Aircraft
Fiber optic sensor with antenna
USAF
Receiving antenna

Fig. 11.49. Health monitoring of structures using fiber optics and antenna technology.

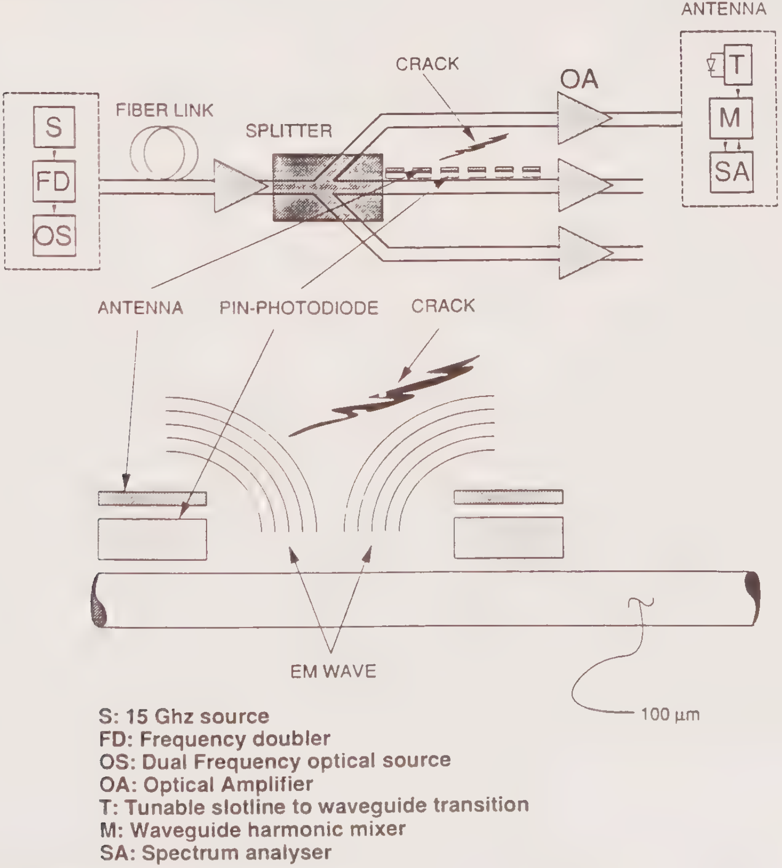

ANTENNA
CRACK
OA
T
M
SA
S
FD
OS
FIBER LINK
SPLITTER
ANTENNA
PIN-PHOTODIODE
CRACK
EM WAVE
100 µm
S: 15 Ghz source
FD: Frequency doubler
OS: Dual Frequency optical source
OA: Optical Amplifier
T: Tunable slotline to waveguide transition
M: Waveguide harmonic mixer
SA: Spectrum analyser

Fig. 11.50. Fiber optic sensor with antenna for 3D imaging of flaws in structures.

Feature selection takes place from the physically based domains of guided waves, MEMS, microwave, etc. The classification will use state-of-the-art artificial neural networks. This multitechnology approach to NDE vastly improves our ability to monitor the health of aircraft, civil structures and manufacturing processes. The goal, of course, is to finally develop a physically based feature vector for a complex system from which data fusion can be used in the development of a decision algorithm for criticality of system performance based on detection, classification, and sizing. The feature vector established with this multimode and multitechnology approach is rich and robust. Rather than filling in the elements of a feature vector with a long list of statistically or probabilistic based features on a single modality NDE approach in an attempt to solve the problem, the approach here with a multimode and multitechnology approach is initially to have tremendous sensitivity potential hence improving the overall probability of detection and decision algorithm performance potential. Once the feature vector is completely formulated, procedures for algorithm development follows traditional guidelines of pattern recognition whereby probability density function curves could be plotted to select the most useful features for final input into a pattern recognition or neural net algorithm, thereby establishing a decision function with appropriate weights for the physically based features selected.

Thus, the multitechnology approach to health monitoring of systems and processes is not simply a set of disparate and complementary technologies, but rather a multiscale inspection ability that can be combined together for a powerful new approach to NDE.

11.8 CONCLUSION

An integration of smart materials, MEMS, and control devices has been presented for aerospace structures. The smart skins derived from this integration are ideally suited for drag sensing and control of aircraft and spacecraft. One can also induce microriblets on the smart skin for drag reduction using micromachining techniques. These microriblets can be actively produced by the MEMS techniques at the desired location. The detection and control of noise and vibration as caused by both external and internal forces may be countered through the use of MEMS sensors and piezoelectric actuators. An array of shear wave sensors introduces a method for localization of the internal stress field throughout the structure. The wireless remote and continuous telemetry systems do not weaken the structure. The wireless telemetry system can be incorporated along with fiber optic sensor for health monitoring of aerospace structures. The advanced polymeric smart materials may also serve as a “smart wall paper” for cabin noise sensing and control of aircraft.

11.9 ACKNOWLEDGMENTS

The work is in part supported by the US Army CECOM, The US Army Research Office, the Office of Naval Research and the industrial members of the Center for the Engineering of Electronic and Acoustic Materials of the Pennsylvania State University. The authors would like to thank Dr. Gary Anderson for many helpful discussions and his support and encouragement. The authors would also like to thank many of their graduate students, in particular, H. Subramanian, D. Piscotty, K. Ramanathan, Y. R. Roh, F. Selmi, J. Kim, and Z. Wu and their colleague Dr. Bao.

REFERENCES

1. S. Middelhoek and S. A. Audet, *Silicon Sensors,* Academic Press, London, 1989.
2. K. E. Peterson, “Silicon as a mechanical material,” Proc. IEEE **70**, 1982, pp. 420–457.
3. J. Bryzek, K. Peterson, and W. McCulley, “Micromachines on the march,” *IEEE Spectrum*, May 1994, pp. 20–31.
4. L.-S. Fan, Y.-C. Tai and R. S. Muller, “Pin joints, gears, springs, cranks and other novel micromechanical structures," Proc. Transducers ’87, pp. 849–852.
5. Proceedings of IEEE MEMS, 1993–1997.
6. K. Ikuta and T. Higuchi, eds., Proceedings of the IEEE Micro Electro Mechanical Systems Piscataway, NJ, 1994.
7. V. K. Varadan, Proc. SPIE 2448, 1995.
8. J. Tani and M. Esashi, eds., Proceedings of the International Symposium on Microsystems, Intelligent Materials and Robots, Tohoku University, Japan, 1995.
9. V. K. Varadan, Proc. SPIE 2722, 1996.
10. K. Ikuta and K. Hirowatari, "Real three dimensional micro fabrication using stereo lithography and metal molding," Proc. IEEE Micro Electro Mechanical Systems, 1993, p. 42.
11. V. K. Varadan and V. V. Varadan, "3D MEMS structures and their applications," International Symposium on Microsystems, Intelligent Materials and Robots, Tohoku University, Japan, 1995.
12. T. Takagi and N. Nakajima, "Photoforming applied to fine machining," Proc. IEEE Micro Electro Mechanical Systems, 1993, p. 173.
13. D. L. Carraway, “The use of silicon microsensors in smart skins for aerodynamic research” Int’l Congress on Instrumentation in Aerospace Simulation Facilities, Dec. 1991, pp. 413–422.
14. L. C. Chin, V. V. Varadan, and V. K. Varadan "Hybrid finite element formulation for periodic piezoelectric arrays subjected to fluid loading," *Int. J. Numerical Methods in Engineering,* Vol. 37, 1994, pp. 2987–3003.

15. H. Aeidel, "The mechanism of anisotropic silicon etching and its relevance for micromachining," *Proc. Int. Conf. Solid-State Sensors and Actuators*, Tokyo, Japan, 1987, pp. 120–125.
16. K. W. Shaw, Z. L. Zhang, and N. C. MacDonald, "SCREAM: a single mask, single-crystal silicon, reactive ion etching process for microelectromechanical structures," *Sensors and Actuators A*, Vol. 40, pp. 63–70, 1994.
17. A. Lakhtakia, V. K. Varadan, and V. V. Varadan, "Highly-directional, steerable, and wide beam launcher," CEEAM Report, 1984.
18. J. H. Smith, S. Montage, and J. J. Sniegowski, "Material and processing issues for the monolithic integration of microelectronics with surface-micromachined polysilicon sensors and actuators," Proc. SPIE 2639, 1995.
19. K. W. Markus, D. A. Koester, A. Cowen, R. Mahadevan, V. R. Dhuler, D. Roberson, and L. Smith, "MEMS infrastructure: the multi-user MEMS processes (MUMPS)," Proc. SPIE 2639, 1995.
20. L. K. Nanver, E. J. G. Goudena, and H. W. van Zeijt, "DIMES-01, a baseline BIFET process for smart sensor experimentation," *Sensors and Actuators A*, Vol. 36, pp. 136–147, 1993.
21. M. J. Vellekoop, G. W. Lubking, P. M. Sarro, and A. Venema, " Integrated-circuit-compatible design and technology of acoutci-wave-based microsensors," *Sensors and Actuators A,* Vol. 44, pp. 249–263, 1994.
22. Herbert Taub and Donald Schilling, *Digital Integrated Electronics,* Mc Graw Hill, Eectrical and Electronic Engineering Series, New Delhi, pp. 331–334, pp. 346–348, 1977.
23. J. P.Huber and M.W. Rosneck, "Successful ASIC design the first time through," Van Nostrand Reinhold, New York, 1991.
24. P. Naish and P. Bishop, *Designing ASICs*, Ellis Horwood Series in Electrical and Electronics Engineering, Halstead Press, New York, pp. 71–90, 1988.
25. W. E. Hord "Microwave and mm wave phase shifters" *Microwave Journal*, Sept. 1989, pp. 95–112.
26. V. K. Varadan, K. A. Jose, V. V. Varadan, R. F. Hughes, and J. F. Kelly, "A novel microwave planar phase shifter," *Microwave Journal*, Vol. 38, pp. 244–254, 1994.
27. V. K. Varadan, D. K. Ghodgaonkar, V. V. Varadan, J. Kelly, and P. Glikerdas, "Ceramic phase shifters for electronically steerable antenna systems," *Microwave Journal*, Vol. 35, pp.116, 1992.
28. V. K. Varadan, K. A. Jose, V. V. Varadan, R. F. Hughes, and J. F. Kelly, "A novel microwave planar phase shifter," *Microwave Journal*, Vol. 38, No. 4, pp. 244–254, 1994.
29. K. A. Jose, R. Hollinger, V. K. Varadan, V. V. Varadan, and J. F. Kelly, "Experimental investigations of electronic beam steering array antennas," *J. Wave-Materials Interaction*, Vol. 8, p. 311, 1993.
30. V. K. Varadan, V. V. Varadan, K. A. Jose, and J. F. Kelly, "Electronically steerable leaky-wave antenna using tunable ferroelectric materials", Journal of Smart Materials and Structures, Vol. 3, pp. 470–475, 1994.

31. V. K. Varadan and V. V. Varadan, "Electronically steerable automobile collision warning antennas using tunable ferroelectric materials," Wireless Symposium, Santa Clara, CA, 1994.
32. V. K. Varadan, F. Selmi, and V. V. Varadan, Voltage Tunable Dielectric Ceramics which Exhibit Low Dielectric Constants and Applications thereof to Electrical Structures, US Patent (Serial No. 08/260,053) awarded 1996.
33. V. K. Varadan and V. V. Varadan, "IDT, SAW and MEMS sensors for measuring deflection, acceleration and ice detection of aircraft," Proc. SPIE 3046, pp.209–219, 1997.
34. X. Q. Bao, V. V. Varadan, and V. K. Varadan, "SAW temperature sensor and remote reading system," *Proc. IEEE Ultrasonic Symposium*, Vol. 1, pp. 583–585, 1987.
35. V. K. Varadan, V. V. Varadan, and X. Q. Bao, "Wireless telemetry systems and their applications," *J. Smart Materials and Structures*, accepted for publication.
36. V. K. Varadan, V. V. Varadan, and X. Q. Bao "Integration of interdigital transducers, MEMS and antennas for smart structures," Proc. SPIE 2722, pp. 95–106, 1996.
37. H. Subramanian, V. V. Varadan, and V. K. Varadan, "Wireless remotely readable microaccelerometer," Proc. SPIE 3046, pp. 220–228, 1997.
38. D.E. Oates, D.L. Symthe, and J.B.Green, " SAW/FET programmable transversal filter with 100 MHz bandwidth and enhanced programmability," 1985 Ultrasonics Symposium.
39. I. Yao and S. A.Reible, "Wide bandwidth acoustoelectric convolvers," 1979 Ultrasonics Symposium.
40. W. Kuehnel, "Modelling of the mechanical behaviour of a differential capacitor acceleration sensor," *Sensors and Actuators*, 1995.
41. R. P.van Kampen, M. J. Vellekoop, P. M. Sarro, and R. F.Wolffenbuttel, "Application of electrostatic feedback to critical damping of an integrated silicon capacitive accelerometer," *Sensons and Actuators*, 1994
42. V. K. Varadan and V. V. Varadan, IDT Microsensors and MEMS and Their Applications, US Patent pending.
43. J. L. Rose, K. M. Rajana, and A. Pilarski, "Guided waves for composite patch repair of aging aircraft," Symposium on Composite Aircraft Repair of Aging Aircraft, Vancouver, B.C., 1995.
44. J.-B. Huang and C.-M. Ho, "Micro riblets for drag reduction," Proc. SPIE 2448, pp. 245–250, 1995.
45. V. K. Varadan and V. V. Varadan, "Drag reduction in aircraft structures," SPIE Conference, San Diego, 1995. This paper is not listed in INCITE.
46. V. V. Varadan and V. K. Varadan, "Microriblets for drag reduction using MEMS technology" SPIE Conference on Smart Materials and Structures, San Diego, 1995
47. M. J. Walsh, *AIAA Journal*, Vol. 21, p. 485, 1983.
48. C. R. Fuller, *J. Sound and Vibration,* Vol. 139, p. 1–15, 1990.
49. R. L. Clark and C. R. Fuller, *J. Acoust. Soc. Am.*, Vol. 92, p. 1521–1533, 1992.

50. V. V. Varadan, V. K. Varadan, and X. Q. Bao, "Smart plates for noise reduction using MEMS technology," Proc. SPIE 2722, pp.230–240, 1996.
51. V. L. Metcalf, C. R. Fuller, R. J. Silcox, and D. E. Brow, *J. Sound and Vibration*, Vol. 153, pp. 387–402, 1992.
52. D. R. Thomas, P. A. Nelson, and S. J. Elliott, *J. Sound and Vibration,* Vol. 167, pp. 113–128, 1993.
53. V. V. Varadan, V. K. Varadan, X. Q. Bao, J. H. Jeng, and C. C. Sung, *J. Sound and Vibration*, Vol. 167, pp. 263–275, 1993.
54. V. V. Varadan, S. J. Tsao, V. K. Varadan, and B. R. Tittman, "Image reconstruction of flaws using ramp response signature," *J. Wave-Material Interaction*, Vol. 10, No. 1, pp.67–78, 1995.
55. V. V. Varadan and J. Chandra, "Mathematics and control in smart structures," SPIE Proceedings on Mathematics and Control of Smart Materials and Structures, San Diego, 1996, 1997.

INDEX

{100} symmetry, 51–53, 78
3-D microstructures, 186–188
3-D patterning, 323
3-D polysilicon structures, 482–486
Absorber side wall tilt, 318
Absorbers, 304
Accelerometers, 392–401
Accelerometers, programmable, 652–660
Acoustic attenuation, 661
Actuators, 357, 408–427, 491, 619–684
Adhesives, 660
Aircraft sensors, 664–684
Amorphous silicon thin-film transistors, 581–587
Analog-to-digital converters, 545
Anisotropic etching, 57–71, 80
Anisotropic etching, mechanisms, 69–71
Atomic steps, 76, 87

Bimetallic actuators, 415
Biological/chemical analysis systems, 16
Biomedical instrumentation, 14–20
Biomedical systems, 518–565
  biocompatibility, 528
  cost, 528
  functionality, 523
  packaging, 555–563
  power considerations, 522
  reliability, 527
  size considerations, 521
Biosensors, 503
Bragg reflector mirrors, 142
Brighteners, 224
Bubble flows, 82
Bubble memory, 252
Buckyballs, 83
Bulk micromachining, 20, 625
Buried metal conductors, 179

Capacitive accelerometers, 394
Capacitive pressure sensors, 389
CdSe thin-film transistors, 592–594
Ceramic microcomponents, 360–367
Ceramic powders, 361
Ceramic shaping techniques, 361
Ceramics, 108
Check valves, 420
Chiral boxes, 87
Circuit modification, 173
CMOS integration, 630
Color filters, 595–598
Comb resonators, 120
Comb transducers, 662
Computer on a chip, 254
Conformal antennas, 630
Conformal line coating, 180
Copper electrodeposition, 274
Cross sectioning, 177
Current density, 214, 227–237

Data detectors, 544
Decorative etching, 178
Deep-etch/shallow diffusion, 116–120
Deep-etch x-ray lithography, 303–325
Defect delineation, 54–56
Deposition, focused ion beam, 164–169
Deuterium isotope separation, 84
Diamond machining, 328–333
Differential pressure flow sensors, 408
Diffusion layer, 231
Direct electrophoretic sol deposition, 362
Displays, 10–12, 491–495
  Fabry-Perot, 454
  liquid crystal, 572–612
Dry etching, 101–152
  of polymers, 101–110

E-beam lithography, 252
Electro-osmosis, 426

Electrochemical model (anisotropic etching), 70
Electrode potential, 201
Electrodeposit properties, 219
Electrodeposit structure, 216
Electrodeposition, 212–226
Electroforming, 242
Electrohydrodynamic micropumps, 427
Electroless copper, 211
Electroless deposition, 208–212
Electroless gold, 212
Electroless nickel, 210
Electrophoresis, 426
Electroplating, 199–287
Electroplating Ni, 105
Electrostatic actuators, 413
Endpoint detection milling technique, 175
Epitaxial growth, 217
Etchants, 57–68, 78–81
Etching, gas-assisted, 160, 178

Fabry-Perot displays, 454
Fabry-Perot interferometers, 439–444
Failure analysis, 176
Fast flow sensors, 402
Field emitter arrays, 131–134
Field emitter tips, sharpening, 127–131
Field emitters, 122–134
Flip chips, 630
Flow sensors, 400–408
Fluorescence radiation 318
Force-balanced accelerometers, 397
Frequency measurement hybrid
technique, 635–642
Fresnel diffraction, 316

Gallium implantation, 191
Gas chemistry, 190
Gas field ion source, 191
Grain orientation, 217

HI-MEMS, 258–264
High aspect ratio processing, 301–377
HNA etching, 45–49, 75–78
Hole filling, 180
Hot embossing, 351
Humidity sensors, 405
Hydration model (anisotropic etching), 69
Hydrogen passivation, 51
Hydroxonium mechanism, 79

Image acquisition, 157
Implantable silicon devices, 530–555
In-line process, 189
Inertial sensors, 12–14
Injection molding, 350
Ink jet print heads, 14
Integrated circuits, application-specific, 633
Interdigital transducers, 646–651, 659
Ion beam resolution, 191
Ion beam system, 156
Ion beams, focused, 155–195
Isotropic etching, 44–57

KOH, 57–65, 68, 78–81

Lasers, dry etching fabrication, 134
Lasers, external cavity, 441
Levelers, 225
LIGA process, 237, 258–264, 280, 301–303, 490
  modified, 321–325
LIGA microstructures, 345–360
Liquid crystal displays
  fabrication, 599
  optical performance, 599–609
  semiconductor materials for, 579–594
  thin-film-transistor-addressed, 572–612
Liquid flow sensors, 405
Liquid metal ion source, 155
Lithography, optical, 273
Lost micromolds, 359

Mach-Zender interferometers, 499
Mask distortion, 318
Mask repair, 169–173
Mask technology (x-ray lithography), 304
Mechanical pumps, 421
Medical monitoring instruments, 19
Membranes, 307

MEMS, 4–39, 624–630
  fabrication, 20–22
  integration, 22–30, 619–684
  optical, 436–504
Metal electrodeposits, examples, 264–270
Micro-optical devices, 435–504
Micro-optical disks, 495
Micro-optics, 354–357
Microaccelerometers, 651–662
Microactuators, 443
Microchemistry, 155–195
Microcomponents, development of, 364
Microdrilling, 333–338
Microencoders, 497
Microfluidics, 354
Microinterferometers, 439–444
Microlenses, 487–490
Micromachined systems for neurophysiological applications, 518–565
Micromachining, five challenges, 82–88
Micromechanical machining, 325–345
Micromechanical optical fiber switches, 456–464
Micromechanical optical switches, 454–456
Micromechanical waveguide switches, 464–474
Micromilling, 338–345
Micromirrors, 487–490
Micromolding of ceramics, 108
Microprobes
  recording, 530–541
  timulating, 548–551
Micropumps, 421–427
Microscope, scanning optical, 497
Microsensors, 645
Microstimulators, muscular, 551–555
Microstrip antennas, 642–645
Microvalves, 410–421
Miller index, 76
Milling (sputtering), 158–162, 175
Minimotors, magnetic, 270–280
Mirrors
  dry etching fabrication, 134–138
  GaAs, 138–142
  silicon, 135–138
Mold micromachining, 21
Molded microparts, 353–360
Molding microstructures, 346–360
Multichip thin film package, 255

Nanomachining, 189
Nerve regeneration electrodes, 541
Neural stimulation/recording instruments, 15
Nonmechanical pumps, 425
Nucleation, 212

Optical choppers, 475–482
Optical elements, 482–486, 490
Optical modulators, 448–454
Optical scanners, 475–482
Organic bases, 66–68, 78
Oxidation-dissolution model (anisotropic etching), 69

Packaging, 555–563
  hermetic, 556–562
  nonhermetic, 562
Pattern transfer, 243
Patterning, 243
Permalloy electrodeposition, 274
Phase shifters, 643
Photo electrons, 316
Piezoelectric actuation, 410
Piezoresistive effect, 383
Piezoresistive pressure sensors, 386
Pits, 78
Plastic molding processes, 345–360
Plated films, stress in, 220
Plating, 199–287
  alloy, 207
  formulations, 223
  single-element, 204

Plating-through-mask technology, 237, 258–264, 280
PMMA resists, 312–315
Pneumatic actuation, 412
Polysilicon thin-film transistors, 587
Polymers, 101–110
  etch characteristics, 101
Preamplifiers, 545

Pressure sensors, 385–391
Prosthetic devices, 20

Radiation chemistry, 311
Reaction molding, 352
Resist development, 311
Resist layer (x-ray lithography), 309
Resist residue, 283
Resonant accelerometers, 399
Resonant pressure sensors, 390

Sacrificial layer technique, 321
Secondary electrons, 319
Sensor classification, 381
Sensors, 357, 381–408
  acoustic, 676
  displacement, 497
  dynamic strain, 668
  fiber optic, 683
  flow, 400–408, 503
  health monitoring, 662
  pressure, 385–391, 502, 674
  shear, 674
  vibration, 671
  waveguide, 499–504
Shape memory alloy external actuators, 413
Silicon devices, implantable, 530–555
Silicon, elastic modulus, 383
Silicon, etching, 43–71
Silicon, expansion coefficient, 383
Silicon field emitters, 122–134
Silicon, mechanical properties of, 382
Silicon microlenses, 135–138
Silicon micromirrors, 135–138
Silicon microstructures, etching of, 111–122
Silicon microstructures, high aspect ratio, 110–122
Silicon, piezoresistive effect, 383
Single-crystal silicon thin-film transistors, 590–592
Single-layer structures, 258
SiO_2 etching, 71–75
Sirtl CrO_3:HF family, 54–56, 75–78
Slip pressing process, 362
Smart electronics, 630–663
Smart materials, 619, 623–624
Smart structures, 619–622, 664–684
Solenoid plunger, 410
Sputtering (milling), 158–162
Stereo lithography, 627
Stress-reducing agents, 225
Surface acoustic wave (SAW) sensors, 646–651, 664–676
Surface micromachining, 21, 626
Surface shear sensors, 405
Surgical tools, 19
Synchrotron radiation, 317

Tactile sensors, 20
Tape casting and stamping, 361
Telemetry systems, wireless, 642, 682
TEM analysis, 181–186
Thermal dilution flow sensors, 407
Thermal transit-time sensors, 407
Thermal two-stage process, 361
Thermopneumatic actuators, 417
Thin films, mechanical properties of, 382
Thin-film transistor arrays, 581–595
Thin-film transistors, 572–612
Transducers, multiple, 524
Tunneling accelerometers, 399
Tunneling structures, 400

Vertical semipermeable membranes, 85
Vias, 180
Voltage regulators, 543

Wafer bonding, 660
Wafer cleaning, 49–51
Waveguides, GaAs, 138–142
Wavelength-tunable photodiodes, 445–448
Wet chemical etching, 41–97
Wet chemical etching, five challenges, 75–82
Whole cell systems, 18

X-ray lithography, 252, 279
X-ray masks, 172

HANDBOOK OF

Microlithography, Micromachining, and Microfabrication

Volume 2: MICROMACHINING AND MICROFABRICATION

P. Rai-Choudhury, Editor

Over the last three decades, microelectronics (integrated circuits) has changed essentially every aspect of our lives, and has been the key defining technology of the twentieth century. The next silicon revolution will be different and more important than simply packing more transistors onto silicon. The power and capability of the chips of tomorrow will transcend simply storing and processing information. The hallmark of this next revolution will be the incorporation of new types of functionality onto the chip—structures that will enable the chip to not only reason, but to sense, act, and communicate. This will result in chips that can affect their surroundings and communicate what they know to the outside world in dramatic ways. The "intelligent" chips of tomorrow will process such things as chemicals, motion, light, and knowledge. These revolutionary developments will have as profound an effect on the next century as the integrated circuit had on the twentieth century.

Micromachining and microfabrication, in combination with microlithography, are rapidly finding applications in many areas, from sensors to actuators to biomedical devices, in addition to their use in microelectronics device manufacturing. Such a dynamic field demands a practical and authoritative handbook for process development and production, and to assist in the training of scientists and engineers. This handbook is written by experts in the field in such a manner so it contains useful process details, recipes, tables, charts, etc., and can be utilized as a reference book or as a textbook. The eleven chapters cover key aspects of micromachining and microfabrication, including numerous device applications:

Micromachining and Trends for the Twenty-First Century • Wet Chemical Etching of Silicon and SiO_2 • Applications of Dry Etching to Microsensors, Field Emitters, and Optical Devices • Focused Ion Beams for Micromachining and Microchemistry • Plating Techniques • High Aspect Ratio Processing • Sensors and Actuators • Micro-optical Devices • Micromachined Systems for Neurophysiological Applications • Thin-Film-Transistor-Addressed Liquid Crystal Displays • Microsensors, Actuators, MEMS, and Electronics for Smart Structures

ISBN 0-85296-911-2
9 780852 969113 >

The Institution of Electrical Engineers
Michael Faraday House
Six Hills Way
Stevenage, Herts., SG1 2AY, UK

KR-133-120